Die Monte-Carlo-Methode und verwandte Fragen

S. M. Ermakow

Ermakow • Die Monte-Carlo-Methode und verwandte Fragen

4138

40.—



S. M. Ermakow
Die Monte-Carlo-Methode



Hochschulbücher für Mathematik

Band 72

Herausgegeben von K. Maruhn und W. Rinow

Die Monte-Carlo-Methode und verwandte Fragen

von S. M. Ermakow

Mit 14 Abbildungen und 5 Tabellen



VEB Deutscher Verlag der Wissenschaften
Berlin 1975

С. М. Ермаков
Метод Монте-Карло и смежные вопросы
Наука, Москва 1971

Übersetzung aus dem Russischen: Prof. Dr. E. Schincke, Dr. M. Schleiff

Verlagslektor: Dipl.-Math. K. Bratz
Umschlaggestaltung: A. Mähler

Printed in the German Democratic Republic
Lizenz-Nr. 206 · 435/100/75
Gesamtherstellung: VEB Druckerei „Thomas Müntzer", 582 Bad Langensalza
LSV 1074
Bestellnummer 570 222 5
EVP 62,— Mark

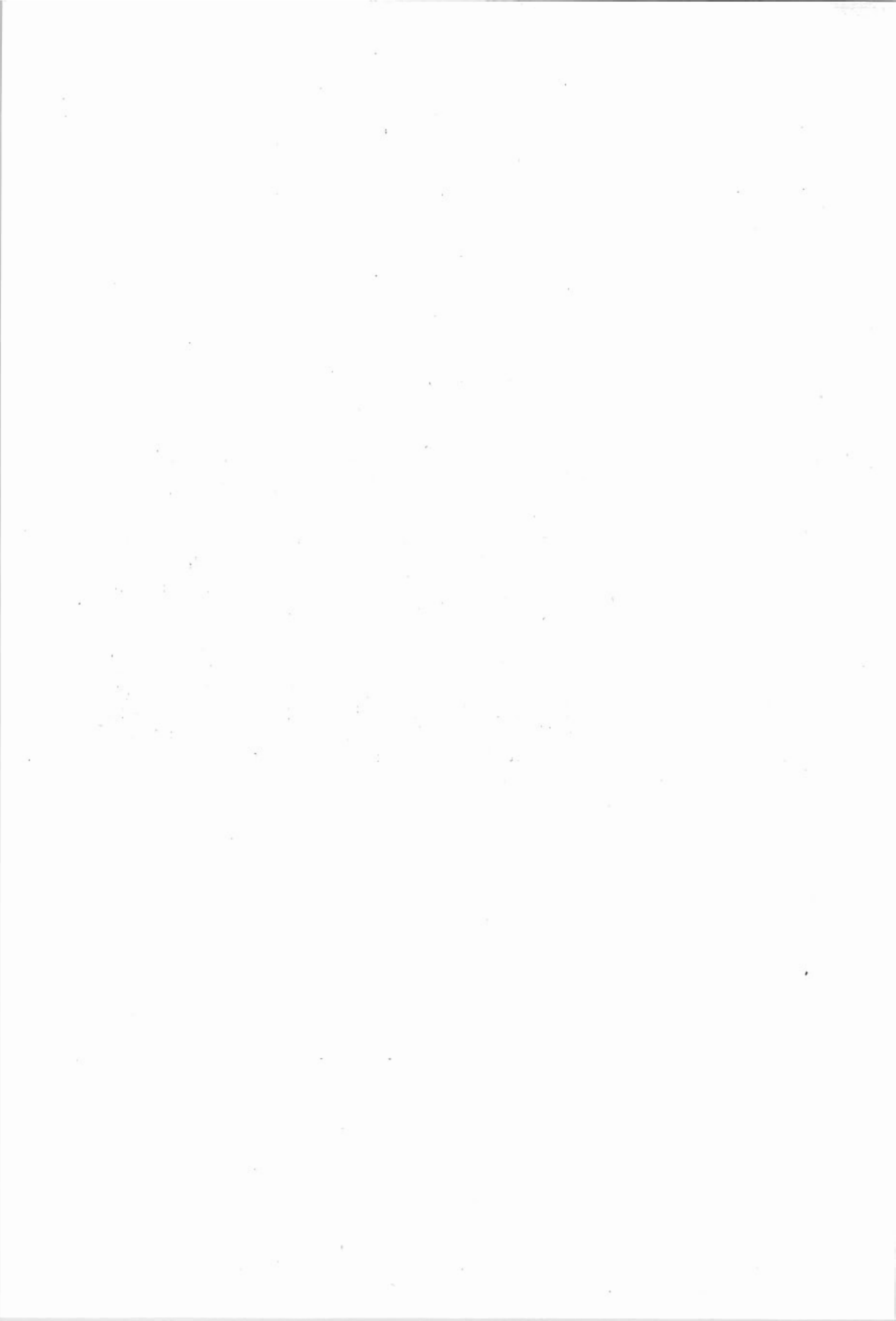
VORWORT ZUR DEUTSCHEN AUSGABE

Es freut mich sehr, daß mein Buch „Die Monte-Carlo-Methode und verwandte Fragen" nun auch einem breiten Kreis deutscher Leser zugänglich wird. Dafür danke ich an dieser Stelle dem VEB Deutscher Verlag der Wissenschaften als Herausgeber und den Übersetzern, Herrn Professor Dr. Erich Schincke und Herrn Dr. Martin Schleiff, Halle. Ich kann mir vorstellen, daß es keine leichte Aufgabe war, diese Übersetzung anzufertigen, zumal die Monte-Carlo-Methode mit vielen angewandten Fragen verknüpft ist, die ihre eigene Terminologie besitzen.

Seit dem Erscheinen der russischen Ausgabe des Buches wurden eine beträchtliche Anzahl von Arbeiten veröffentlicht, die zur Weiterentwicklung der Monte-Carlo-Methode und ihrer Anwendungen wesentlich beitrugen. Das zeugt sowohl von der raschen Entwicklung der Methode selbst als auch von der Aktualität einer Reihe theoretischer Aufgaben, die in das Buch aufgenommen wurden.

In der vorliegenden Ausgabe wurden einige Korrekturen vorgenommen und eine Liste mit Titeln hinzugefügt, die nach 1965 in der ganzen Welt erschienen sind. Bei einer solchen Bibliographie ist es schwer, Anspruch auf Vollständigkeit zu erheben, aber ich hoffe, daß die grundlegenden Arbeiten theoretischen Charakters in dieser Bibliographie enthalten sind. Dabei ist zu beachten, daß in die Ergänzungsliste solche Titel nicht aufgenommen wurden, die schon in den Bibliographien der Comp. Reviews auftreten

S. M. Ermakow



VORWORT DER ÜBERSETZER

Die unter dem Begriff Monte-Carlo-Methode zusammengefaßten Verfahren der numerischen Mathematik haben in den vergangenen Jahren in immer größerem Umfang bei der Lösung verschiedenartiger numerischer Probleme Anwendung gefunden. Während bereits einige in die Monte-Carlo-Methode einführende Bücher in deutscher Sprache vorliegen, fehlte bisher ein Buch, das die Vielfalt der Anwendungsmöglichkeiten dieser Methode von einem fundierten mathematischen Standpunkt her darstellt. Das im Jahr 1971 in Moskau erschienene Buch von Prof. Dr. S. M. Ermakow bietet gerade in dieser Hinsicht eine ausführliche, aber doch kompakte Darstellung der Verfahren der Monte-Carlo-Methode. Die Herausgeber der deutschen Übersetzung hoffen daher, mit dem vorliegenden Buch eine bislang vorhandene Lücke schließen zu können.

An den Universitäten wird gewöhnlich die Monte-Carlo-Methode in Spezialvorlesungen für Studenten höherer Semester behandelt. Das vorliegende Buch bietet eine gute Ergänzung zu einer derartigen Vorlesung; insbesondere wird das grundlegende Problem der Integralberechnung in großer Ausführlichkeit unter verschiedensten Gesichtspunkten dargestellt. Aber auch für ein einführendes Selbststudium ist das Buch durchaus geeignet, da alle in der Monte-Carlo-Methode benötigten Grundkenntnisse der Wahrscheinlichkeitstheorie und Statistik am Anfang kurz dargestellt werden. Dem in der Praxis tätigen Mathematiker bietet dieses Buch ebenfalls viele Anregungen. Der Autor war bemüht, nicht bei einer rein theoretischen Beschreibung der Verfahren der Monte-Carlo-Methode stehenzubleiben, sondern stets eine fundierte Einschätzung der Effektivität der einzelnen numerischen Verfahren zu geben. Zur Erhöhung der Effektivität werden hauptsächlich verschiedene Methoden zur Dispersionsverkleinerung untersucht. Der Mathematiker findet hier also viele Hinweise, wann die Anwendung der Monte-Carlo-Methode sinnvoll ist und in welcher Weise diese Anwendung erfolgen sollte.

Schließlich bietet das Buch auch dem in der Forschung tätigen Mathematiker zahlreiche Anregungen. Der Autor weist an vielen Stellen auf noch offene Probleme hin und gibt damit den Anstoß zur weiteren Untersuchung der mit der Monte-Carlo-Methode verbundenen Fragestellungen.

Bei der Übersetzung wurde angestrebt, eine möglichst wortgetreue Wiedergabe zu erreichen, damit die im russischen Original bewährte Darstellung erhalten bleibt. Andererseits wurde an manchen Stellen auch eine freiere Übertragung gewählt, wenn damit das Verständnis erleichtert und eine größere Einheitlichkeit erreicht wurde. Im russischen Original waren eine größere Zahl von Druckfehlern, aber auch einige Ungenauigkeiten enthalten. Die Übersetzer

waren bemüht, diese Unstimmigkeiten zu korrigieren, ohne daß darauf im einzelnen hingewiesen wird. Eine wesentliche Hilfe war dabei ein vom Autor zur Verfügung gestelltes Fehlerverzeichnis mit entsprechenden Berichtigungen.

Für die Reinschrift des Manuskripts danken die Herausgeber Frau E. Wolf.
Schließlich sei dem Verlag und der Druckerei für die ansprechende Gestaltung der deutschen Ausgabe gedankt.

Erich Schincke
Martin Schleiff

VORWORT ZUR RUSSISCHEN AUSGABE

Dieses Buch stellt eine erweiterte Fassung der Vorlesungen dar, die der Autor an der mathematisch-mechanischen Fakultät der Leningrader staatlichen Universität gehalten hat.

Bekanntlich setzt sich die Monte-Carlo-Methode aus drei Bestandteilen zusammen: erstens die Erzeugung von Zufallsgrößen mit gegebenem Verteilungsgesetz, zweitens der Aufbau stochastischer Modelle realer Prozesse (Systeme) und drittens die Lösung von gewissen Aufgaben aus der statistischen Theorie der Schätzungen. Die unmittelbare Modellierung realer Prozesse führt bei weitem nicht immer zum Erfolg, und in diesen Fällen entsteht die Aufgabe, gewisse ,,fiktive" Modelle zu schaffen, die eine effektivere Lösung der gestellten Aufgabe gestatten.

Diese letzte Aufgabe und auch spezielle Aufgaben aus der statistischen Theorie der Schätzungen, die bei der Monte-Carlo-Methode auftreten, werden in diesem Buch bevorzugt behandelt.

Die ersten drei Kapitel des Buches stellen im wesentlichen eine Einführung dar, um einen Einblick in die Anwendbarkeit und die numerischen Besonderheiten der Methode zu geben. Einen zentralen Platz nimmt das vierte Kapitel ein, in dem Methoden zur Integralberechnung betrachtet werden. In den Kapiteln V und VI deren Inhalt sich unmittelbar an die Ausführungen von Kapitel IV anschließt, und in den weiteren Kapiteln bis einschließlich Kapitel VII wird von der Möglichkeit Gebrauch gemacht, eine Folge von Realisierungen einer gleichverteilten Zufallsgröße zu finden. Erst im siebenten Kapitel wird diese Frage ausführlicher erörtert.

Es ist natürlich, daß die speziellen Interessen des Verfassers Einfluß auf die Stoffauswahl hatten. Das Buch ist daher keinesfalls eine erschöpfende Abhandlung über die Monte-Carlo-Methode. Eine Reihe von Fragen wird nur flüchtig berührt, und ein Teil des Stoffes hat mehr den Charakter einer Übersicht. Der Autor hofft aber, daß Fragen im Zusammenhang mit der Rolle der statistischen Theorie der Schätzungen und die Aufgabe der Dispersionsverkleinerung genügend vollständig behandelt wurden, obwohl die schnelle Entwicklung von Arbeiten in dieser Richtung die letzte Behauptung etwas einschränkt. So können die Materialien der 2. Allunionskonferenz zu Monte-Carlo-Methoden, die stattfand, als das Buch im wesentlichen schon fertig geschrieben war, als nützliche Ergänzung zu einzelnen seiner Abschnitte dienen.

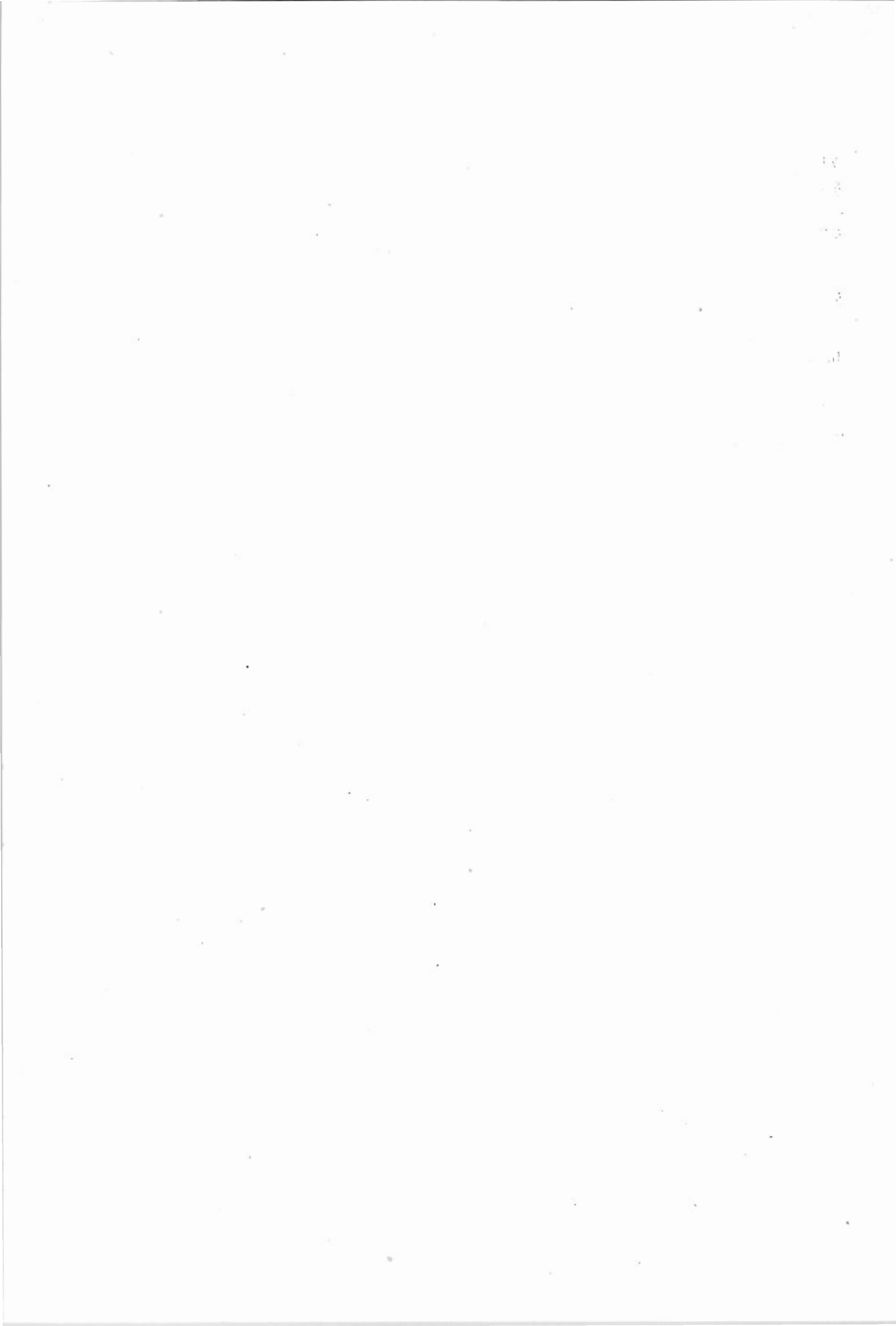
Der § 4 des vierten Kapitels wurde gemeinsam mit B. L. Granowski verfaßt. Diskussionen mit G. A. Michailow, I. M. Sobol, N. N. Tschenzow haben

wesentlichen Einfluß auf die Bearbeitung vieler Probleme der Monte-Carlo-Methode durch den Autor gehabt.

Der Autor ist den genannten Personen und im Besonderen N. N. Tschenzow, dessen Bemerkungen zur Beseitigung einer Reihe von Fehlern im Manuskript führten, zutiefst dankbar.

Der Autor dankt auch J. W. Linnik und G. I. Martschuk, die den Autor während seiner Ausbildung auf den Themenkreis gelenkt haben, der in diesem Buch zur Darstellung gelangt.

INHALT



I. EINFÜHRUNG

§ 1. Einführende und historische Bemerkungen

Die Monte-Carlo-Methode kann definiert werden als Methode zur Modellierung von Zufallsgrößen mit dem Ziel, spezielle Eigenschaften ihrer Verteilungen zu berechnen. In der Regel wird angenommen, daß die Modellierung auf elektronischen Datenverarbeitungsanlagen (EDVA) verwirklicht wird, obwohl man in einigen Fällen schon bei der Benutzung solcher Hilfsmittel wie Roulettespiel, Bleistift und Papier Erfolg haben kann.

Die erste Idee zur Ausnutzung zufälliger Erscheinungen auf dem Gebiet der numerischen Rechnungen stammt aus dem Jahre 1878, als die Arbeit von Hall [1] über die Bestimmung der Zahl π durch zufälliges Werfen einer Nadel auf parallel liniertes Papier erschien. Diese Aufgabe ist allgemein bekannt, und ihre Beschreibung kann man in Büchern der Wahrscheinlichkeitstheorie finden (z. B. bei Prochorow und Rosanow [1], S. 17). Das Wesen der Sache besteht darin, daß man experimentell Ergebnisse nachbildet, deren Wahrscheinlichkeit durch die Zahl π ausgedrückt wird, und daß man diese Wahrscheinlichkeit näherungsweise abschätzen kann. Obwohl sich analoge Überlegungen in bezug auf andere Aufgaben in einer Reihe späterer Arbeiten finden (vgl. dazu Hammersley und Handscomb [1]), enthalten sie doch noch keine bemerkenswerte Entwicklung. Erst im Jahre 1944, als im Zusammenhang mit Arbeiten an dem Atombombenprojekt John von Neumann vorschlug, den Apparat der Wahrscheinlichkeitstheorie für die Lösung angewandter Aufgaben mit Hilfe der EDV auszunutzen, begann die eigentliche Entwicklung der Monte-Carlo-Methode. Die erste Arbeit, die solche Probleme systematisch darstellte, ist die von Metropolis und Ulam [1].

Arbeiten sowjetischer Wissenschaftler zur Monte-Carlo-Methode erscheinen seit 1955/56. Innerhalb von zwei Jahrzehnten entstand eine umfangreiche Bibliographie zur Monte-Carlo-Methode, die mehr als 2000 Titel umfaßt. Dabei zeigt bereits ein flüchtiger Blick auf die Titel der Arbeiten die Brauchbarkeit der Monte-Carlo-Methode zur Lösung angewandter Aufgaben aus verschiedenen Gebieten von Wissenschaft und Technik.

Ursprünglich wurde die Monte-Carlo-Methode hauptsächlich zur Lösung von Aufgaben der Neutronenphysik benutzt, bei denen die traditionellen numerischen Methoden wenig brauchbar sind. Weiter erstreckte sich ihr Einfluß auf eine große Klasse von Aufgaben der statistischen Physik, die ihrem Inhalt nach sehr verschieden sind. Zu Wissenschaftszweigen, die in besonders großem Maß die Monte-Carlo-Methode benutzen, muß man u. a. die Bedienungstheorie, die Spieltheorie und mathematische Ökonomie und die Theorie der Nachrichtenübertragung beim Vorhandensein von Störungen zählen.

Die Monte-Carlo-Methode übte bisher und übt noch jetzt einen wesentlichen Einfluß auf die Entwicklung von Methoden der numerischen Mathematik aus (z. B. Entwicklung von numerischen Integrationsverfahren). Bei der Lösung vieler Aufgaben läßt sie sich erfolgreich mit anderen numerischen Verfahren verbinden und ergänzt diese. Ihre Anwendung ist in erster Linie bei solchen Aufgaben gerechtfertigt, die eine wahrscheinlichkeitstheoretische Beschreibung gestatten. Das ist erklärlich, da die Antwort bei Aufgaben mit stochastischem Inhalt mit einer gewissen Wahrscheinlichkeit gegeben wird und sich außerdem die Lösungsprozedur bei der Monte-Carlo-Methode wesentlich vereinfacht. Die Schwierigkeit bei der Lösung dieser oder jener Aufgabe auf einer EDVA entsteht in bedeutendem Maß durch Schwierigkeiten bei der Übersetzung in die Maschinensprache. Die Schaffung der Programmiersprachen vereinfachte eine Etappe dieser Arbeit wesentlich. Die kompliziertesten Etappen sind daher gegenwärtig die mathematische Beschreibung der zu untersuchenden Erscheinung, die notwendige Vereinfachung der Aufgabe, die Auswahl der geeigneten numerischen Methode, die Untersuchung der dabei auftretenden Fehler und das Aufstellen des Algorithmus. In den Fällen, in denen es eine wahrscheinlichkeitstheoretische Beschreibung der Aufgabe gibt, kann die Benutzung der Monte-Carlo-Methode die genannten einzelnen Etappen wesentlich vereinfachen. Wie sich im Weiteren herausstellen wird, ist es übrigens in vielen Fällen nützlich, auch für streng deterministische Aufgaben ein stochastisches Modell zu konstruieren (randomisieren der Aufgabe), damit dann die Monte-Carlo-Methode benutzt werden kann.

Obgleich eine Definition der Monte-Carlo-Methode hier gegeben wurde, wird der Inhalt dieser Methode, wie gewöhnlich, erst durch die weiteren Ausführungen völlig klar. Wir verwenden hier immer die Bezeichnung „Monte-Carlo-Methode“, die mit der Bezeichnung „Methode der statistischen Versuche“ gleichbedeutend ist, und auch in der sowjetischen Literatur gebräuchlich ist. In der nichtsowjetischen Literatur spricht man gewöhnlich dann von der Monte-Carlo-Methode, wenn man von der Vorstellung ausgeht, daß die der Berechnung zugrunde liegenden Größen von verschiedenen stochastischen Modellen ausgehend geschätzt werden können (z. B. können Zufallsgrößen mit verschiedenem Verteilungsgesetz den gleichen Mittelwert haben). Im folgenden bietet es sich aber an, von verschiedenen Modellen und ihren Transformationen im Rahmen einer Methode zu sprechen. Oft, wenn vom Studium irgendeiner realen Erscheinung die Rede ist, bezeichnet man die Modellierung durch Zufallsgrößen als „Simulation“ (russ. имитация).

Wir werden die Bezeichnung „Simulation“ manchmal auch gebrauchen, um die Verbindung des entsprechenden Modells mit gewissen realen Erscheinungen zu unterstreichen, obwohl die Schaffung des stochastischen Modells einer Erscheinung notwendigerweise der Modellierung vorangeht, und es erhebt sich dann die Frage nach einer mehr oder weniger präzisen Beschreibung.

§ 2. Einige Definitionen und Fakten aus der Wahrscheinlichkeitstheorie

In diesem Paragraphen und in dem gesamten Kapitel werden die weiterhin notwendigen Hilfsmittel bereitgestellt und wird eine Einführung in die Wahrschein-

lichkeitstheorie und die mathematische Statistik gegeben. Es sind verhältnismäßig wenig Begriffe und Fakten aus diesen Gebieten, die im folgenden benötigt werden, weil man vollkommen ohne den analytischen Apparat der Wahrscheinlichkeitstheorie auskommt. Die grundlegenden Definitionen werden zur Bequemlichkeit des Lesers anschließend formuliert. Deren ausführliche Darstellung würde das Buch unvermeidlich in eine Abhandlung über Wahrscheinlichkeitstheorie und mathematische Statistik verwandeln und vom eigentlichen Ziel ablenken. Daher werden nur solche Fakten ohne Beweis formuliert, die große Bedeutung für das im Weiteren Dargestellte haben oder verhältnismäßig selten in der üblichen Literatur anzutreffen sind.

1. Wahrscheinlichkeitsräume. Wie üblich gehen wir von einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathsf{P})$ aus, wobei

1. Ω eine gegebene Menge ist, genannt *Raum der Elementarereignisse*; seine Elemente werden auch als Punkte bezeichnet;

2. $\mathfrak{A}$ eine nichtleere Menge von Untermengen von Ω ist, die eine *σ-Algebra* darstellt, d. h. abgeschlossen ist bezüglich der Summe (Vereinigung) abzählbar vieler Mengen sowie des Durchschnitts und der Komplementbildung; vorausgesetzt, daß Ω (und folglich auch die leere Menge) Elemente von $\mathfrak{A}$ sind;

3. P ein *Wahrscheinlichkeitsmaß* ist, d. h. eine nichtnegative Funktion über $\mathfrak{A}$ mit $\mathsf{P}(\Omega) = 1$ und

$$\mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) ,$$

wenn die A_i eine Folge von paarweise durchschnittsfremden Mengen aus $\mathfrak{A}$ sind. Außer den Wahrscheinlichkeitsmaßen werden oft auch *endliche Maße* $\tilde{\mu}$ betrachtet, die sich vom Wahrscheinlichkeitsmaß durch eine andere Normierung $\tilde{\mu}(\Omega) = \text{const} < \infty$ unterscheiden, und *σ-endliche Maße* μ, für die $\mu(\Omega)$ unendlich sein kann, aber der ganze Raum Ω eine Darstellung als Summe abzählbar vieler Mengen A_i mit endlichem Maß $\mu(A_i) < \infty$, $i = 1, 2, \ldots$, gestattet.

Das übliche Vorgehen bei Definitionen eines Maßes auf einer σ-Algebra besteht darin, daß das Maß zunächst nur auf einer engeren Gesamtheit von Mengen definiert wird, und zwar auf einem Halbring, und danach auf die den Halbring umfassende σ-Algebra ausgedehnt wird. Eine Gesamtheit $\boldsymbol{A}$ von Mengen heißt *Halbring*, wenn mit zwei beliebigen Mengen aus $\boldsymbol{A}$ auch der Durchschnitt zu $\boldsymbol{A}$ gehört. Dabei kann eine beliebige Menge aus $\boldsymbol{A}$ als Summe von höchstens abzählbar vielen durchschnittsfremden Mengen $A_1, A_2, A_3, \ldots$ dargestellt werden. Insbesondere ist ganz Ω darstellbar in der Form $\Omega = \bigcup_i A_i$. Es existiert eine minimale σ-Algebra, die den gegebenen Halbring $\boldsymbol{A}$ umfaßt, d. h. eine solche, die im Durchschnitt aller $\boldsymbol{A}$ umfassenden σ-Algebren enthalten ist.

Ist $A \in \mathfrak{A}$ eine beliebige Menge aus $\mathfrak{A}$ und gilt $A \subseteq \bigcup_i A_i$, wobei alle A_i aus $\boldsymbol{A}$ sind, so erhält man durch

$$\mu(A) = \inf \sum_i \mu(A_i)$$

eine Fortsetzung des Maßes μ vom Halbring $\boldsymbol{A}$ auf die σ-Algebra $\mathfrak{A}$. Dabei ist das Infimum bezüglich jeder möglichen Auswahl von abzählbar vielen Mengen A_i aus $\boldsymbol{A}$ zu bestimmen. Eine solche Fortsetzung eines σ-endlichen Maßes auf die σ-Algebra $\mathfrak{A}$ ist in eindeutiger Weise möglich.

Als einfachstes Beispiel eines Halbrings kann die Gesamtheit aller halboffenen Intervalle $(x', x'']$ der reellen Achse dienen. Die minimale σ-Algebra, die den Halbring dieser Intervalle umfaßt, trägt den Namen σ-Algebra der Borelschen Mengen der Zahlengeraden. Das Maß auf dem Halbring der Intervalle kann durch die Gleichung

$$\mathsf{P}\big((x', x'']\big) = F(x'') - F(x')$$

gegeben sein, wobei $F(x)$ eine beliebige, auf $-\infty < x < +\infty$ definierte, nichtnegative, rechtsseitig stetige, monoton nichtfallende Funktion ist mit

$$\lim_{x\to-\infty} F(x) = 0 \qquad \text{und} \qquad \lim_{x\to+\infty} F(x) = 1 .$$

Jede Funktion mit den aufgezählten Eigenschaften definiert ein Wahrscheinlichkeitsmaß auf den Borelschen Mengen der Zahlengeraden. Weiter wird häufig als Raum Ω der s-dimensionale euklidische Raum (s kann auch unendlich sein) oder ein linearer normierter Funktionenraum vorkommen, und es wird die durch den Halbring der offenen Menge dieser Räume erzeugte σ-Algebra betrachtet Eine solche σ-Algebra trägt auch die Bezeichnung *σ-Algebra der Borelschen Mengen*, und ihre Elemente heißen *Borelsche Mengen*.

Sind $(\Omega_1, \mathfrak{A}_1, \mathsf{P}_1)$ und $(\Omega_2, \mathfrak{A}_2, \mathsf{P}_2)$ zwei Wahrscheinlichkeitsräume, so ist das direkte (kartesische) Produkt $\Omega_1 \times \Omega_2$ die Menge aller geordneten Paare (ω_1, ω_2), $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$. Wir definieren das Produkt der σ-Algebren

$$\mathfrak{A}^{(1,2)} = \mathfrak{A}_1 \times \mathfrak{A}_2$$

als die kleinste σ-Algebra, die den Halbring aller möglichen Rechteckmengen der Art $A_1 \times A_2$ $(A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2)$ umfaßt. Das Maß auf diesem Halbring kann durch

$$\mathsf{P}(A_1, A_2) = \mathsf{P}_1(A_1) \cdot \mathsf{P}_2(A_2)$$

definiert und auf die σ-Algebra $\mathfrak{A}^{(1,2)}$ erweitert werden. Ein auf diese Weise bestimmtes Maß $\mathsf{P}_1 \times \mathsf{P}_2$ trägt die Bezeichnung *Produkt der Maße* P_1 und P_2. Der Wahrscheinlichkeitsraum $(\Omega_1 \times \Omega_2, \mathfrak{A}_1 \times \mathfrak{A}_2, \mathsf{P}_1 \times \mathsf{P}_2)$ wird Produkt der Räume $(\Omega_1, \mathfrak{A}_1, \mathsf{P}_1)$ und $(\Omega_2, \mathfrak{A}_2, \mathsf{P}_2)$ genannt. Das Produkt einer beliebigen Anzahl von Wahrscheinlichkeitsräumen kann dann durch Induktion definiert werden.

2. Meßbare Funktionen, Integral, Ableitung. Nach Einführung eines Maßes auf einer σ-Algebra ist es ohne weiteres möglich, den Begriff des Integrals einer meßbaren reellen Funktion auf dem Raum Ω zu definieren. Das geschieht völlig analog zur Definition des Lebesgueschen Integrals.

Die Funktion $\xi(\omega)$ $(\omega \in \Omega)$ heißt $\mathfrak{A}$-*meßbar* (oder *einfach meßbar*, wenn es klar ist, welche σ-Algebra gemeint ist), wenn die Menge $\{\omega : \xi(\omega) < x\}$ für jede reelle Zahl x zu $\mathfrak{A}$ gehört. Eine meßbare Funktion heißt *einfach*, wenn sie auf Ω nicht mehr als abzählbar viele verschiedene Werte annimmt. Wenn die Menge $\{\omega : \xi(\omega) < x\}$ für alle x eine Borelmenge ist, was wir in der Regel voraussetzen, so heißt $\xi(\omega)$ eine *Borelfunktion*. Die Klasse der Borelfunktionen fällt zusammen mit der

Klasse der Baireschen Funktionen, falls Ω ein linearer normierter Raum R ist. Dabei enthält die Klasse der Baireschen Funktionen alle Funktionen auf R, die sich als Grenzwert einer bezüglich der Metrik von R konvergenten Folge stetiger Funktionen darstellen lassen.

Nun sei A_i eine Folge durchschnittsfremder Mengen, auf denen die einfache Funktion $\eta(\omega)$ konstante Werte η_i hat. Man kann zeigen, daß diese Mengen meßbar sind. $\eta(\omega)$ heißt *integrierbar*, wenn die Reihe $\sum_{i=1}^{\infty} |\eta_i| \mathsf{P}(A_i)$ konvergiert, und die Größe

$$\int_{\Omega} \eta(\omega)\, \mathsf{P}(d\omega) = \sum_{i=1}^{\infty} \eta_i \mathsf{P}(A_i)$$

heißt *Integral* der einfachen Funktion η über dem Raum Ω. Eine beliebige meßbare Funktion $\xi(\omega)$ kann für fast alle ω als Grenzwert einer konvergenten Folge einfacher Funktionen dargestellt werden, d. h., mit Ausnahme einer Menge vom Maß Null gilt

$$\xi(\omega) = \lim_{n\to\infty} \eta_n(\omega) \qquad \big(\omega \in \Omega \ (\mathrm{mod}\ \mathsf{P})\big).$$

Hier und weiterhin wird die Bezeichnung (mod P) gebraucht, wenn eine Eigenschaft für alle Elemente einer Menge mit Ausnahme einer Menge vom P-Maß Null erfüllt ist (P Wahrscheinlichkeitsmaß oder σ-endliches Maß). In diesem Fall sagen wir gewöhnlich, daß die Eigenschaft für fast alle ω bezüglich des Maßes P gilt oder mit der Wahrscheinlichkeit 1 gilt, wenn P ein *Wahrscheinlichkeitsmaß* ist.

Alle Funktionen, für die die obere Grenze

$$\sup_{\eta} \int_{\Omega} |\eta(\omega)|\, \mathsf{P}(d\omega)$$

bezüglich aller einfachen Funktionen η mit $|\eta| < |\xi|$ endlich ist, heißen *integrierbar*. Das Integral $\int_{\Omega} \xi(\omega)\, \mathsf{P}(d\omega)$ wird als Grenzwert von Integralen solcher einfacher Funktionen η_n definiert, daß $\eta_n(\omega)$ für fast alle ω für $n \to \infty$ gegen $\xi(\omega)$ strebt und $|\eta_n| \leqq |\xi|$ gilt. Das Integral über eine beliebige Menge A aus $\mathfrak{A}$ existiert, wenn die Funktion $\xi\varphi_A$ integrierbar ist. Dabei ist φ_A die *charakteristische Funktion der Menge* A, also

$$\varphi_A(\omega) = \begin{cases} 1 & \text{für} \quad \omega \in A\,, \\ 0 & \text{für} \quad \omega \notin A\,. \end{cases}$$

Dann ist

$$\int_{A} \xi(\omega)\, \mathsf{P}(d\omega) = \int_{\Omega} \xi(\omega)\, \varphi_A(\omega)\, \mathsf{P}(d\omega)\,.$$

Für das auf diese Weise definierte Integral sind die üblichen Eigenschaften des Lebesgueschen Integrals über ein endliches Intervall gültig, was wir im weiteren ohne nähere Begründung unter Hinweis auf entsprechende Literatur benutzen werden.

Wenn ein σ-endliches Maß $\mu(A)$ $(A \in \mathfrak{A})$ die Darstellung

$$\mu(A) = \int_{A} f(\omega)\, \mathsf{P}(d\omega) \tag{1.2.1}$$

gestattet, wobei f eine nichtnegative integrierbare Funktion ist, so heißt $f(\omega)$ die *Radon-Nikodymsche Ableitung* des Maßes μ nach dem Maß P, und man gebraucht die Schreibweise

$$f(\omega) = \frac{d\mu}{d\mathsf{P}}(\omega) . \tag{1.2.2}$$

Nach dem Satz von RADON-NIKODYM existiert die Ableitung $f(\omega)$ für gegebene σ-endliche Maße μ und P, d. h., es gibt eine Darstellung (1.2.1) genau dann, wenn μ absolut stetig in bezug auf P ist ($\mu \ll \mathsf{P}$), d. h., aus der Gleichung $\mathsf{P}(A) = 0$, $(A \in \mathfrak{A})$ folgt $\mu(A) = 0$.

3. Zufallsgrößen und Verteilungsfunktionen. Es sei $(\Omega, \mathfrak{A}, \mathsf{P})$ ein Wahrscheinlichkeitsraum. Die Mengen A aus $\mathfrak{A}$ heißen *Ereignisse* und die Größen $\mathsf{P}(A)$ ihre *Wahrscheinlichkeiten.*

Eine beliebige $\mathfrak{A}$-meßbare reelle Funktion $\xi(\omega)$, die fast überall endliche Werte annimmt, heißt eine auf Ω definierte *Zufallsgröße.* Zufallsgrößen, die sich nur auf einer Menge vom Maß Null unterscheiden, werden als gleich betrachtet.

Für jede Zufallsgröße $\xi(\omega)$ ist eindeutig die *Verteilungsfunktion*

$$F_\xi(x) = \mathsf{P}(\omega : \xi(\omega) < x) \tag{1.2.3}$$

definiert, die eine monoton nichtfallende, nichtnegative, linksseitig stetige Funktion mit

$$\lim_{x \to -\infty} F_\xi(x) = 0 , \qquad \lim_{x \to +\infty} F_\xi(x) = 1$$

darstellt. Wie früher gezeigt wurde, wird mit Hilfe von $F_\xi(x)$ eindeutig das Wahrscheinlichkeitsmaß P auf den Borelmengen der Zahlengeraden bestimmt. Dieses Maß nennt man auch die Verteilung der Zufallsgröße $\xi(\omega)$. Wenn $\Xi = (\xi_1, ..., \xi_s)$ eine Gesamtheit von auf $(\Omega, \mathfrak{A}, \mathsf{P})$ definierten Zufallsgrößen, also einen Zufallsvektor, darstellt, so heißt die Funktion

$$F_\Xi(x_1, ..., x_s) = \mathsf{P}\left(\bigcap_{i=1}^{s} \omega : \xi_i(\omega) < x_i\right)$$

gemeinsame Verteilungsfunktion der Zufallsgrößen $\xi_1, ..., \xi_s$ oder *Verteilungsfunktion des Zufallsvektors* Ξ. Sie ist eindeutig bestimmt im s-dimensionalen euklidischen Raum, monoton nicht fallend und linksseitig stetig in jeder Variablen. Außerdem gilt

$$\lim_{x_k \to -\infty} F_\Xi(x_1, ..., x_s) = 0 , \qquad k = 1, ..., s ,$$

und

$$\lim_{x_{i+1}, ..., x_s \to \infty} F_\Xi(x_1, ..., x_i, ..., x_s) = F_{\Xi_i}(x_1, ..., x_i)$$

mit

$$\Xi_i = (\xi_1, ..., \xi_s) ,$$

so daß $F_{\Xi_i}(x_1, ..., x_i)$ eine Verteilungsfunktion des Zufallsvektors Ξ_i darstellt. Mit Hilfe der Funktion $F_\Xi(x_1, ..., x_s)$ kann unter einigen zusätzlichen Bedingungen (vgl. PROCHOROW und ROSANOW [1]) auch das Maß auf den Borelmengen des s-dimensionalen euklidischen Raumes definiert werden.

Wenn das Integral

$$\int_{\Omega} \xi(\omega)\,\mathsf{P}(d\omega) = \int_{-\infty}^{\infty} x\, dF_{\xi}(x)$$

existiert, so nennt man dessen Wert den *Erwartungswert* (*mathematische Erwartung*) der Zufallsgröße ξ und bezeichnet ihn mit $\mathsf{M}\xi$. Wenn das Wahrscheinlichkeitsmaß P absolut stetig in bezug auf das σ-endliche Maß μ ist, so nennt man die Radon-Nikodymsche Ableitung $\frac{d\mathsf{P}}{d\mu}(\omega)$ die *Verteilungsdichte* von P bezüglich μ. Wenn insbesondere μ das Lebesguesche Maß ist und fast überall die gewöhnliche Ableitung $\frac{dF_{\xi}(x)}{dx} = p_{\xi}(x)$ existiert, so heißt diese einfach die *Verteilungsdichte* oder *Wahrscheinlichkeitsdichte* der Zufallsgröße ξ. Für mehrdimensionale Zufallsgrößen wird die Dichte der gemeinsamen Verteilung durch die Gleichung

$$p_{\Xi}(x_1, \ldots, x_s) = \frac{\partial^s F_{\Xi}(x_1, \ldots, x_s)}{\partial x_1 \cdots \partial x_s}$$

für fast alle $x_1, \ldots, x_s$ definiert.

Mit Hilfe des früher eingeführten Begriffs des Integrals über einer Menge A aus $\mathfrak{A}$ können *die bedingte Wahrscheinlichkeit* und *der bedingte Erwartungswert* definiert werden. Es sei $\mathfrak{B}$ eine in $\mathfrak{A}$ enthaltene σ-Algebra. Bedingter Erwartungswert $\mathsf{M}(\xi|\mathfrak{B})$ der Zufallsgröße ξ bezüglich $\mathfrak{B}$ heißt jede $\mathfrak{B}$-meßbare Funktion, die für beliebige B aus $\mathfrak{B}$ der Gleichung

$$\int_{B} \xi(\omega)\,\mathsf{P}(d\omega) = \int_{B} \mathsf{M}(\xi|\mathfrak{B})\,\mathsf{P}(d\omega)$$

genügt. Unter Benutzung des Satzes von Radon-Nikodym kann man zeigen, daß eine solche Funktion bis auf eine Menge vom Maß Null immer existiert und eindeutig bestimmt ist. Als bedingte Wahrscheinlichkeit $\mathsf{P}(B|\mathfrak{B})$ eines Ereignisses $B \in \mathfrak{B}$ bezüglich der σ-Algebra $\mathfrak{B}$ wird die Funktion $\mathsf{M}(\varphi_B|\mathfrak{B})$, wobei φ_B die charakteristische Funktion von B ist, bezeichnet. Diese ist $\mathfrak{B}$-meßbar und für beliebige $A \in \mathfrak{B}$ gilt, mit Ausnahme einer Menge vom Maß Null, die Gleichung

$$\mathsf{P}(A \cap B) = \int_{A} \mathsf{P}(B|\mathfrak{B})\,\mathsf{P}(d\omega)\,,$$

die auch zur Definition der bedingten Wahrscheinlichkeit dienen kann. In dem von uns betrachteten Fall, daß die Zufallsgröße reelle Werte aus $(-\infty, +\infty)$ annimmt, gibt es unter den die bedingte Wahrscheinlichkeit definierenden äquivalenten Funktionenmengen eine solche, die Eigenschaften wie ein Wahrscheinlichkeitsmaß hat. Das erlaubt es (mit der Wahrscheinlichkeit 1), eine bedingte Verteilungsfunktion zu definieren:

$$F_{\xi}(x|\mathfrak{B}) = \mathsf{P}(\omega : \xi < x|\mathfrak{B})\,.$$

Es sei U eine Menge mit einer bestimmten σ-Algebra ihrer Teilmengen $\mathfrak{C}$ (meßbarer Raum), und die Funktion g soll eine meßbare Abbildung $\Omega \to U$ definieren. Als bedingter Erwartungswert (bzw. bedingte Wahrscheinlichkeit) bezüglich der Funktion g wird der bedingte Erwartungswert (bzw. die bedingte Wahrschein-

lichkeit) in bezug auf die σ-Algebra $g^{-1}\mathfrak{C}$ bezeichnet. Analog wird für eine Familie von Funktionen g_i $(i = 1, \ldots, n)$ mit Werten in meßbaren Räumen $\mathfrak{C}_i$ der bedingte Erwartungswert $\mathsf{M}(\xi | g_1, \ldots, g_n)$ definiert als bedingter Erwartungswert bezüglich der minimalen σ-Algebra, die alle $g_i^{-1}\mathfrak{C}_i$ umfaßt. Wie oben kann der Begriff der bedingten Verteilungsfunktion bezüglich einer gegebenen Funktion oder eines Funktionensystems eingeführt werden.

4. Momente von Zufallsgrößen, das Gesetz der großen Zahlen und der zentrale Grenzwertsatz. Existiert die Größe $\mathsf{M}\xi^r$, $r > 0$, so heißt sie r-tes Moment der Zufallsgröße ξ. Die Größe $\mathsf{M}(\xi - \mathsf{M}\xi)^r$ heißt r-tes zentrales Moment. Das zweite zentrale Moment $\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2$ wird Dispersion genannt.

Eine wichtige Rolle werden im weiteren die zweiten zentralen Momente der Komponenten des Zufallsvektors spielen, die *Dispersionen*

$$\mathsf{D}\xi_k = \mathsf{M}(\xi_k - \mathsf{M}\xi_k)^2$$

und die Kovarianz

$$\operatorname{cov}(\xi_k, \xi_l) = \mathsf{M}[(\xi_k - \mathsf{M}\xi_k)(\xi_l - \mathsf{M}\xi_l)], \qquad k \neq l,$$
$$\operatorname{cov}(\xi_l, \xi_l) = \mathsf{D}\xi_l .$$

Die Matrix $\mathcal{D} = ||\operatorname{cov}(\xi_k, \xi_l)||_{k,l=1,\ldots,s}$ nennt man *Kovarianzmatrix*.

Mit den zweiten Momenten einer Zufallsgröße ist die Normalverteilung eng verbunden. Eine Zufallsgröße ξ heißt *normalverteilt* auf $(-\infty, \infty)$, wenn sie die Verteilungsfunktion

$$\Phi\left(\frac{x-a}{\sigma}\right) = (\sqrt{2\pi}\sigma)^{-1} \int_{-\infty}^{x} e^{-(x-a)^2/2\sigma^2}\, dx \tag{1.2.4}$$

besitzt, wobei $a = \mathsf{M}\xi$ und $\sigma^2 = \mathsf{D}\xi$ ist. Der Zufallsvektor $\Xi = (\xi_1, \ldots, \xi_s)$ heißt im s-dimensionalen Raum *normalverteilt*, wenn für seine Dichte

$$p_\Xi(x_1, \ldots, x_s) = (2\pi)^{-s/2} (\det \mathcal{D})^{-1/2} \exp\left(-\tfrac{1}{2}\, \tilde{\mathfrak{x}}^\mathsf{T} \mathcal{D}^{-1} \tilde{\mathfrak{x}}\right) \tag{1.2.5}$$

gilt. Dabei ist $\tilde{\mathfrak{x}}$ ein Spaltenvektor mit den Komponenten $x_1 - a_1, \ldots, x_s - a_s$, $a_k = \mathsf{M}\xi_k$, und $\mathcal{D}$ ist die Kovarianzmatrix des Vektors $\tilde{\mathfrak{x}}$.

Die Möglichkeit, eine Aufgabe mit der Monte-Carlo-Methode zu lösen, findet gewöhnlich ihre Begründung im *Gesetz der großen Zahlen* für unabhängige Zufallsgrößen. Es seien $\xi_1, \ldots, \xi_n$ unabhängige Zufallsgrößen mit gleicher Verteilung. Wir nehmen an, daß

$$\mathsf{M}\xi_k = a < \infty \qquad \text{und} \qquad \mathsf{D}\xi_k = \sigma^2 < \infty \qquad (k = 1, \ldots, n)$$

existieren, und setzen $S_n = \sum_{k=1}^{n} \xi_k$. Für beliebiges $\varepsilon > 0$ gilt dann die *Tschebyscheffsche Ungleichung*

$$\mathsf{P}\left(\left|\frac{S_n}{n} - a\right| \leqq \varepsilon\right) \geqq 1 - \frac{\sigma^2}{n\varepsilon^2} \tag{1.2.6}$$

oder anders formuliert

$$\mathsf{P}\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) \leqq \frac{\sigma^2}{n\varepsilon^2}. \tag{1.2.7}$$

Daraus folgt unter den genannten Voraussetzungen die einfachste Variante des Gesetzes der großen Zahlen. Für beliebiges $\varepsilon > 0$ und $\delta > 0$ und genügend großes n unterscheidet sich das arithmetische Mittel S_n/n mit einer Wahrscheinlichkeit die nicht kleiner als $1 - \delta$ ist, von a um höchstens den Wert ε. Tatsächlich gilt das Gesetz der großen Zahlen noch unter wesentlich schwächeren Voraussetzungen. Insbesondere behauptet das Gesetz der großen Zahlen in der Form von Chintschin (Feller [1]), daß

$$\mathsf{P}\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right)_{n\to\infty} \to 0$$

gilt, falls nur die Zufallsgröße ξ einen endlichen Erwartungswert besitzt. Daher konvergiert das arithmetische Mittel S_n/n unter sehr allgemeinen Voraussetzungen der Wahrscheinlichkeit nach gegen den Erwartungswert. Ist die Dispersion $\mathsf{D}\xi_k$ unendlich, so verlieren die Ungleichungen (1.2.6) ihren Sinn, aber es gilt noch die Ungleichung (von Bahr und Esseen [1])

$$\mathsf{M}\left|\frac{S_n}{n} - a\right|^r \leqq \frac{2}{n^r} \sum_{k=1}^{n} \mathsf{M}|\xi_k - a|^r, \quad 1 \leqq r \leqq 2\,, \tag{1.2.8}$$

wenn die ξ_k unabhängige Zufallsgrößen mit nicht notwendig identischer Verteilung, aber dem gleichen Erwartungswert a sind. Im allgemeinen Fall kann die Konstante in (1.2.8) nicht verbessert werden. Schließlich werden wir im weiteren das starke Gesetz der großen Zahlen benutzen, in dem behauptet wird, daß das arithmetische Mittel S_n/n unabhängiger identisch verteilter Zufallsgrößen gegen ihren Erwartungswert mit der Wahrscheinlichkeit 1 konvergiert.

Für uns ist nicht nur die Konvergenz der Größe S_n/n gegen den Erwartungswert von Interesse, sondern auch die Verteilung von S_n/n. Ist $\sigma^2 < \infty$ und existiert das dritte Moment $\beta_3 = \mathsf{M}|\xi_k - a|^3 < \infty$, so besagt der zentrale Grenzwertsatz, daß S_n/n asymptotisch normalverteilt ist mit der Dispersion σ^2/n und dem Erwartungswert a. Es gilt dabei die Ungleichung

$$\left|\mathsf{P}(S_n - an > \sigma\sqrt{n}x) - \frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-u^2/2}\,du\right| \leqq \frac{c_0\beta_3}{\sigma^3\sqrt{n}}\,, \tag{1.2.9}$$

wobei c_0 eine positive Konstante mit der (bis jetzt bekannten) Abschätzung $1/\sqrt{2\pi} \leqq c_0 < 0{,}9051$ ist. Ibragimow und Linnik [1] haben hinreichende und notwendige Bedingungen dafür angegeben, daß die Konvergenz gegen die Normalverteilung von der Ordnung $o(n^{-1/2})$ erfolgt. Im allgemeinen Fall sind die Grenzverteilungen des arithmetischen Mittels unabhängiger Zufallsgrößen sogenannte stabile Verteilungen, wie sie ausführlich von Gnedenko und Kolmogoroff [1] betrachtet wurden. Wir werden nicht auf ihre Eigenschaften eingehen, weil sie bis jetzt verhältnismäßig selten in der Monte-Carlo-Methode Anwendung finden.

5. Zufällige Prozesse. Der Begriff des Zufallsprozesses (stochastischer Prozeß) ist eine unmittelbare Verallgemeinerung des Begriffs Zufallsgröße. Es sei ϑ ein Parameter, der Werte aus der Menge Θ annimmt. Die Funktion $\xi(\vartheta, \omega)$, die als Funktion von ϑ auf Θ und als Funktion von ω auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathsf{P})$ definiert ist, ist für jeden festen Wert ϑ aus Θ eine Zufallsgröße $\xi(\omega)$ und wird *Zufallsprozeß* genannt. Gewöhnlich wird vorausgesetzt, daß Θ die Menge der reellen Zahlen ist. Ist Θ eine Menge komplizierterer Art, so heißt $\xi(\vartheta, \omega)$ eine *Zufallsfunktion* oder ein *Zufallsfeld.*

Alle möglichen gemeinsamen Verteilungen der Zufallsgrößen $\xi(\vartheta_1, \omega), \dots, \xi(\vartheta_n,\omega)$ nennt man endlichdimensionale Verteilungen des Zufallsprozesses.

Nach dem Satz von Kolmogoroff [1] existiert für jede Menge von Funktionen $F_n(x_1, \vartheta_1; \dots; x_n, \vartheta_n)$, die bei beliebigen festen n und $\vartheta_1, \dots, \vartheta_n$ als Verteilungsfunktionen erscheinen, ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathsf{P})$ und eine Familie meßbarer Funktionen $\xi(\vartheta, \omega)$, $\omega \in \Omega$, derart, daß

$$F_n(x_1, \vartheta_1; \dots; x_n, \vartheta_n) = \mathsf{P}\big(\xi(\vartheta_1, \omega) \leqq x_1; \dots; \xi(\vartheta_n, \omega) \leqq x_n\big)$$

gilt, wenn die Verträglichkeitsbedingung

$$\lim_{x_n \to \infty} F_n(x_1, \vartheta_1; \dots; x_n, \vartheta_n) = F_{n-1}(x_1, \vartheta_1; \dots; x_{n-1}, \vartheta_{n-1})$$

erfüllt ist. Auf diese Weise kann ein Zufallsprozeß durch eine Funktion F_n für alle möglichen n definiert sein.

Eine Funktion des Parameters ϑ, die für jedes feste ϑ gleich $\mathsf{M}\xi(\vartheta, \omega)$ ist, heißt *Erwartungswert* oder *Mittelwert* des Zufallsprozesses, und eine Funktion zweier Veränderlicher

$$R(\vartheta, t) = \mathsf{M}\big(\xi(\vartheta, \omega)\, \xi(t, \omega)\big)$$

heißt *Kovarianzfunktion.* Unten werden wichtige Sonderfälle von Zufallsprozessen betrachtet. Die Prozesse werden dabei im engeren Sinn definiert (vgl. Doob [1]).

Ein Zufallsprozeß $\xi(\vartheta) = \xi(\vartheta, \omega)$ heißt *stationär,* wenn die mit ihm verbundene Wahrscheinlichkeitsverteilung invariant gegenüber Verschiebungen des Parameters ϑ ist. Sind alle endlichdimensionalen Verteilungen des Prozesses Normalverteilungen, so spricht man von einem *Gaußschen Prozeß.*

Ein Prozeß heißt *Markoffscher Prozeß,* wenn für seine bedingten endlichdimensionalen Verteilungen mit der Wahrscheinlichkeit 1 die Gleichungen

$$\mathsf{P}\big(\xi(\vartheta_n) \leqq x | \xi(\vartheta_1), \dots, \xi(\vartheta_{n-1})\big) = \mathsf{P}\big(\xi(\vartheta_n) \leqq x | \xi(\vartheta_{n-1})\big) \tag{1.2.10}$$

bei beliebigen Parameterwerten $\vartheta_1 \leqq \cdots \leqq \vartheta_n$ aus Θ erfüllt sind.

Die Bedingung (1.2.10) ist der folgenden äquivalent: Für beliebige $t_1 < t_2$ aus Θ und jedes x gilt mit der Wahrscheinlichkeit 1

$$\mathsf{P}\big(\xi(t_2) \leqq x | \xi(t), t \leqq t_1\big) = \mathsf{P}\big(\xi(t_2) \leqq x | \xi(t_1)\big)\,. \tag{1.2.11}$$

Ein Markoffscher Prozeß ist also gegeben durch die Übergangswahrscheinlichkeit und die Anfangsverteilung. Es wird in diesem Fall vorausgesetzt, daß Θ ein minimales Element ϑ_0 hat, für das die Anfangsverteilung π der Zufallsgröße $\xi(\vartheta_0)$ gegeben ist, und daß die Funktion (Übergangswahrscheinlichkeit) $\mathsf{P}(\omega, t; \omega', \vartheta)$;

$\omega, \omega' \in \Omega$, $t, \vartheta \in \Theta$, für fast alle ω der Gleichung von CHAPMAN-KOLMOGOROFF

$$\mathsf{P}(\omega, t; \omega', \vartheta) = \int \mathsf{P}(\omega'', \tau; \omega', \vartheta)\, \mathsf{P}(\omega, t; d\omega'', \tau)$$

genügt. Diese Gleichung sichert, daß die Verträglichkeitsbedingungen der endlichdimensionalen Verteilungen des Prozesses erfüllt sind.

Es sei ξ ein stationärer Markoffscher Prozeß, also $\mathsf{P}(\omega, t; \omega', \vartheta) = \mathsf{P}(\omega, \omega')$, und φ_B die charakteristische Funktion der Borelschen Menge $B \in \mathfrak{A}$. Für beliebige $\vartheta_0 < \vartheta_1 < \cdots < \vartheta_n$ wird die gemeinsame Verteilung der Zufallsgrößen $\xi_0 = \xi(\vartheta_0, \omega), \ldots, \xi_n = \xi(\vartheta_n, \omega)$ durch den Ausdruck

$$\int \pi(d\xi_0) \int \mathsf{P}(\xi_0, d\xi_1) \cdots \int \varphi_B \mathsf{P}(\xi_{n-1}, d\xi_n) \tag{1.2.12}$$

definiert.

Ein wichtiger Sonderfall des Markoffschen Prozesses ist die Markoffsche Kette mit einer endlichen Anzahl von Zuständen m. In diesem Fall kann die Funktion $\xi(\vartheta, \omega)$ nur endlich viele Werte annehmen. Im stationären Fall wird die Kette für $\vartheta = \vartheta_0$ durch eine diskrete Anfangsverteilung beschrieben. Die Rolle der Übergangswahrscheinlichkeit spielt hier die Übergangsmatrix (stochastische Matrix

$$\mathcal{P} = ||p_{i,j}||_{i,j=1,\ldots,m},$$

deren Elemente den Bedingungen $p_{i,j} \geqq 0$, $\sum_{j=1}^{m} p_{i,j} = 1$ genügen.

Im wesentlichen werden wir in den folgenden Kapiteln nur Markoffsche Prozesse oder sogar nur ihren Sonderfall, die Markoffschen Ketten, mit endlich oder abzählbar unendlich vielen Zuständen betrachten. Als einzige Ausnahme erscheint der (eindimensionale) Prozeß der Brownschen Bewegung, der eine unmittelbare Beziehung zur Lösung der Wärmeleitungsgleichung hat. Der Prozeß der Brownschen Bewegung, der ein Gaußscher Prozeß ist, ist eng verknüpft mit dem Wienerschen Maß und dem Wienerschen Integral, deren exakte Einführung recht kompliziert ist. Wir beschränken uns daher nur auf eine Erläuterung dieser Begriffe. Die ausführliche formale Darstellung findet der Leser in einer Reihe von Lehrbüchern (z. B. SCHILOW und FAN DYK TINH [1]). In dem Buch von KAC [1] ist eine ausgezeichnete Darstellung der Anwendungen enthalten, die das Wienersche Maß und Integral in der Physik finden.

Wie schon aus der Bezeichnung folgt, ist der Prozeß der Brownschen Bewegung eng mit der Brownschen Bewegung von Teilchen verbunden. Wenn ein Teilchen die Bewegung zur Zeit $t = 0$ im Punkt $x = 0$ beginnt, so ist die Wahrscheinlichkeit dafür, daß es sich zu den Zeitpunkten $0 < t_1 < t_2 < \cdots < t_n$ jeweils in den Intervallen $[a_1, b_1], \ldots, [a_n, b_n]$ befindet, durch die Formel

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} p(0, x_1; t_1)\, p(x_1, x_2; t_2 - t_1) \cdots p(x_{n-1}, x_n; t_n - t_{n-1})\, dx_1 \cdots dx_n \tag{1.2.13}$$

mit

$$p(x, y; t) = (2\pi t)^{-1/2} \exp\left[-\frac{1}{2t}(y - x)^2\right] \tag{1.2.14}$$

gegeben. Wurde als Raum Ω der Raum der auf $[0, \infty)$ stetigen Funktionen $x(t)$ mit $x_i = x(t_i)$ für $i = 1, \ldots, n$ und $x(0) = 0$ gewählt, so definiert die Formel

(1.2.13) ein Wahrscheinlichkeitsmaß auf dem Halbring der Mengen von der Form

$$[a_1 \leqq x(t_1) \leqq b_1; \ldots; a_n \leqq x(t_n) \leqq b_n]$$

(es ist leicht zu verifizieren, daß die Verträglichkeitsbedingungen erfüllt sind), und dieses kann auf den ganzen Raum Ω der stetigen Funktionen fortgesetzt werden. Entsprechend dem allgemeinen Schema bei der Einführung des Integrals kann nun weiter das Wienersche Integral definiert werden.

Es sei $f(x)$ eine stetige nichtnegative Funktion über $x \in (-\infty, +\infty)$. Das Wienersche Integral

$$J = \int \exp\left[-\int_0^t f(x(\tau))\, d\tau\right] \mathsf{P}(d\omega) \tag{1.2.15}$$

mit dem Wienerschen Maß P kann auch als Grenzwert

$$J = \lim_{n\to\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{t}{n}\sum_{i=1}^{n} f(x_i)\right] p(0, x_1; t_1) \cdots p(x_{n-1}, x_n; t_n - t_{n-1})\, dx_1 \cdots dx_n$$

betrachtet werden, den man mittels (1.2.14) folgendermaßen schreiben kann:

$$J = \lim_{n\to\infty} (2\pi t n^{-1})^{-n/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[-t n^{-1}\left\{\sum_{i=1}^{n} f(x_i) + \frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - x_{i-1})^2}{t n^{-1}}\right\}\right] dx_1 \cdots dx_n.$$

Das erlaubt, J in der Form

$$\int \exp\left[-\int_0^t \left\{\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2 + f(x(\tau))\right\} d\tau\right] d \text{ (Trajektorie)}$$

zu schreiben. Im folgenden wird uns eine solche symbolische Form der Schreibweise eines Integrals nützlich sein. Auf diese Weise werden im Fall des Prozesses der Brownschen Bewegung die Punkte des gewählten Raumes Ω mit den stetigen Funktionen identifiziert.

Die Funktion $\xi(\vartheta, \omega)$ heißt bei festem ω *Realisierung des Zufallsprozesses*. Die Realisierungen eines Zufallsprozesses können als Elemente eines Funktionenraumes betrachtet werden. Unter bestimmten Bedingungen kann auch eine Metrik im Funktionenraum der Trajektorien eingeführt werden. Für die Feststellung der Stetigkeit der Trajektorien des Prozesses wird das Stetigkeitskriterium für Trajektorien von Kolmogoroff für separable zufällige Prozesse benutzt. Wenn Θ der reellen Achse angehört und $\mathfrak{A}$ ein gewisses System von Borelschen Mengen ist, so heißt der Prozeß ξ *separabel bezüglich* $\mathfrak{A}$, wenn es eine Folge $\vartheta_1, \vartheta_2, \ldots$ von Parameterwerten und eine Menge $\Omega_1 \in \Omega$ vom P-Maß Null derart gibt, daß sich mit $A \in \mathfrak{A}$ und einem beliebigen offenen Intervall I die Mengen

$$\{\omega : \xi(\vartheta) \in A, \vartheta \in I\} \qquad \text{und} \qquad \{\omega : \xi(\vartheta_j) \in A, \vartheta_j \in I\}$$

untereinander nur um eine Teilmenge von Ω_1 unterscheiden. Insbesondere ist der Prozeß der Brownschen Bewegung separabel.

Das Kriterium für die Stetigkeit der Trajektorien behauptet nun folgendes:

Es sei $\xi(\vartheta)$, $\vartheta \in [a, b]$, ein separabler Zufallsprozeß. Wenn für gewisse $\beta > 0$, $\varepsilon > 0$ und $0 < C < \infty$ und alle hinreichend kleinen $\Delta\vartheta$ die Ungleichung

$$\mathsf{M}|\xi(\vartheta) - \xi(\vartheta + \Delta\vartheta)|^\beta \leqq C|\Delta\vartheta|^{1+\varepsilon}$$

gilt, so sind fast alle Realisierungen ξ des Prozesses stetige Funktionen von ϑ.

Es seien weiter $\xi(t_i, \omega) = \xi_i$, $i = 1, 2, \ldots$, die Realisierungen des Zufallsprozesses für $t = t_i$. Die Folge von Werten ξ_i wird stationär genannt, wenn der zugrundeliegende Prozeß stationär ist. Die Zufallsgrößen ξ_i bilden abhängige Zufallsgrößen. Unter gewissen Voraussetzungen über die schwache Abhängigkeit der Zufallsgrößen, die die stationäre Folge bilden, ist die Grenzverteilung des arithmetischen Mittels der ξ_i stabil. Eine ausführliche Theorie zu dieser Frage kann man in der Monographie von Ibragimow und Linnik [1] finden, in der auch Bedingungen formuliert werden, unter denen die Grenzverteilung eine Normalverteilung ist.

Wir beschränken uns bei der Formulierung des Grenzwertsatzes nur auf den Sonderfall der k-abhängigen Zufallsgrößen. Die Zufallsgrößen $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$ einer stationären Folge heißen *k-abhängig*, wenn zufällige Vektoren der Art $(\xi_{q-r}, \xi_{q-r+1}, \ldots, \xi_q)$ und $(\xi_p, \xi_{p+1}, \ldots, \xi_{p+s})$ unabhängig sind, falls nur $p - q > k$ ist. Setzt man

$$\sigma^2 = \mathsf{M}(\xi_0^2) + 2 \sum_{j=1}^{\infty} \mathsf{M}\xi_0\xi_j ,$$

dann gilt im Fall $0 < \sigma^2 < \infty$

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} \xi_j < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du . \tag{1.2.16}$$

Der entsprechende Grenzwertsatz für Markoffsche Ketten wird im sechsten Kapitel formuliert. Grenzwertsätze für abhängige Zufallsgrößen werden häufig in der Monte-Carlo-Methode benutzt.

§ 3. Einige Erläuterungen zur mathematischen Statistik und das allgemeine Schema der Monte-Carlo-Methode

1. Schätzung der Parameter von Verteilungen. Wie schon bemerkt, liegt in der Monte-Carlo-Methode der Konstruktion einer kompliziert verteilten Zufallsgröße eine auf dem Intervall [0, 1] gleichverteilte Zufallsgröße α zugrunde. Mit anderen Worten, der zugrunde liegende Wahrscheinlichkeitsraum soll konstruktiv sein.

Die Idee einer konstruktiven Aufgabenstellung geht zurück auf die Schöpfer der Theorie der Zufallsprozesse N. Wiener und P. Lévy. In seinem Buch „Zufallsprozesse und Brownsche Bewegung" fordert P. Lévy [1], daß ein Prozeß modelliert werden kann, indem man von einer abzählbaren Folge von Zufallsexperimenten ausgeht, und er beschreibt eine Konstruktion solcher Art. Bereits früher wurde die Modellierung der Brownschen Bewegung von N. Wiener behandelt (vgl. Paley und Wiener [1]). Die entsprechenden Algorithmen werden im zweiten Kapitel angeführt.

Im Anschluß daran wurde die Idee der konstruktiven Aufgabenstellung bei Zufallsprozessen in einer Reihe von Arbeiten, z. B. von GELFAND, FROLOW und TSCHENZOW [1], ausgenutzt und weiterentwickelt. Die arithmetische Modellierung von Zufallsprozessen wird kurz im siebenten Kapitel betrachtet.

Unter einem konstruktiven Wahrscheinlichkeitsraum werden wir nun einen solchen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathsf{P})$ verstehen, für den eine meßbare Abbildung $\omega = \varphi(x)$, $x \in [0, 1]$, des Segments $[0, 1]$ in Ω derart existiert, daß für jedes $A \in \mathfrak{A}$

$$\boldsymbol{\mu}\{A\} = \int_B dx \qquad \text{mit} \qquad B = \{x : \varphi(x) \in A\} .$$

gilt (vgl. TSCHENZOW [2]).

Alle im folgenden zu betrachtenden Wahrscheinlichkeitsräume werden als konstruktiv vorausgesetzt. Methoden zur Konstruktion der Abbildung $\omega = \varphi(x)$, d. h. Methoden, mit deren Hilfe die Realisierungen von Zufallsgrößen mit verschiedenen Verteilungsgesetzen durch die Realisierung von α ausgedrückt werden, werden ausführlich im zweiten Kapitel untersucht. In vielen konkreten Fällen sind sie in numerischer Hinsicht aber recht kompliziert.

Nachdem eine Reihe von Realisierungen der uns interessierenden Zufallsgröße ξ vorliegt, besteht die Aufgabe der Monte-Carlo-Methode nun darin, Kenntnisse über ihre Verteilung zu erhalten. Das ist eine typische Aufgabe der mathematischen Statistik, die Methoden zur Parameterschätzung der Verteilung einer Zufallsgröße auf Grund ihrer Realisierungen untersucht. Den Terminus Parameter einer Verteilung werden wir genügend weit verstehen, so daß insbesondere Funktionale der Art $\int_\Omega g(\xi)\, \mathsf{P}(d\omega) = \vartheta$ mit beliebiger P-meßbarer Funktion $g(\xi)$ eingeschlossen sind. Für die Schätzung eines Parameters ϑ durch Stichprobenwerte (Realisierungen) $\xi_1, \ldots, \xi_N$ der Zufallsgröße ξ wird eine gewisse Funktion $\hat{\vartheta}(\xi_1, \ldots, \xi_N)$ (Statistik) konstruiert, die man Parameterschätzung nennt. Dabei werden die $\xi_1, \ldots, \xi_N$ als Zufallsgrößen betrachtet, genauer als Zufallsvektor, der auf einem Wahrscheinlichkeitsraum definiert ist, dessen Konstruktion von der Art der Realisierung der Größen $\xi_1, \ldots, \xi_N$ abhängt.

Wenn $(\Omega, \mathfrak{A}, \mathsf{P})$ der Wahrscheinlichkeitsraum ist, auf dem ξ definiert ist, und $\xi_1, \ldots, \xi_N$ unabhängige Realisierungen (wiederholte Stichproben) von ξ sind, ist der Stichprobenraum $(\Omega^N, \mathfrak{A}^N, \mathsf{P}^N)$, auf dem $\Xi = (\xi_1, \ldots, \xi_N)$ definiert ist, das Produkt von N Exemplaren des ursprünglichen Wahrscheinlichkeitsraumes, also

$$\Omega^N = \Omega \times \cdots \times \Omega \qquad (N \text{ Faktoren}),$$

$\mathfrak{A}^N$ ist die σ-Algebra von Mengen der Art

$$A^N = A_1 \times A_2 \times \cdots \times A_N , \qquad A_j \in \mathfrak{A} ,$$

und

$$\mathsf{P}_\Xi(A_1, \ldots, A_N) = \mathsf{P}_{\xi_1}(A_1) \cdots \mathsf{P}_{\xi_N}(A_N) .$$

Nun sei ϑ irgendein Parameter aus dem Parameterraum Θ. Man nennt eine Schätzung $\hat{\vartheta}(\xi_1, \ldots, \xi_N)$, die eine meßbare Funktion auf $(\Omega^N, \mathfrak{A}^N, \mathsf{P}^N)$ darstellt, *erwartungstreue* oder *unverzerrte Schätzung* von ϑ, wenn der Erwartungswert von $\hat{\vartheta}$ existiert und $\mathsf{M}\hat{\vartheta}(\xi_1, \ldots, \xi_N) = \vartheta$ gilt. Man sagt, daß $\hat{\vartheta}$ *asymptotisch erwartungs-*

treu ist, wenn $\mathsf{M}\hat{\vartheta}(\xi_1, \dots, \xi_N) \to \vartheta$ für $N \to \infty$ gilt. Konvergiert $\vartheta(\xi_1, \dots, \xi_N)$ der Wahrscheinlichkeit nach gegen ϑ für $N \to \infty$, so heißt die Schätzung $\hat{\vartheta}$ (schwach) *konsistent*. Ist $\hat{\vartheta}$ eine erwartungstreue Schätzung mit endlicher Dispersion und existiert dabei keine andere erwartungstreue Schätzung mit kleinerer Dispersion, so nennt man $\hat{\vartheta}$ eine *effektive* Schätzung von ϑ. In dem Fall gibt die Dispersion der Schätzung den Grad ihrer Güte an. Im allgemeinen Fall können andere Kriterien eingeführt werden. Ausführlicher werden wir uns mit dieser Frage im Zusammenhang mit konkreten Aufgaben beschäftigen.

Es sei jetzt ϑ ein Zahlenparameter, und $\bar{\vartheta}(\xi_1, \dots, \xi_N)$, $\underline{\vartheta}(\xi_1, \dots, \xi_N)$ seien zwei reellwertige Funktionen der Stichprobenwerte mit $\underline{\vartheta} < \bar{\vartheta}$. Wenn für irgendein $0 \leqq \gamma \leqq 1$ die Funktionen $\underline{\vartheta}$ und $\bar{\vartheta}$ so ausgewählt werden können, daß $\mathsf{P}(\underline{\vartheta} < \vartheta < \bar{\vartheta}) = \gamma$ gilt, nennt man $(\underline{\vartheta}, \bar{\vartheta})$ ein *Vertrauensintervall* (Konfidenzintervall) für ϑ, das dem Konfidenzniveau γ entspricht (oder 100γ-prozentiges Vertrauensintervall).

Wir nehmen nun an, daß die Dichte der gemeinsamen Verteilung der Stichprobe $\xi_1, \dots, \xi_N$ existiert, und schreiben dafür $p_\Xi(x_1, \dots, x_N; \vartheta)$. Es ist üblich, diese Dichte *Likelihood-Funktion der Stichprobe* in bezug auf ϑ zu nennen. Existiert eine solche Funktion $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_N)$, so daß $p_\Xi(x_1, \dots, x_N; \vartheta)$ seinen maximalen Wert für $\vartheta = \hat{\vartheta}$ annimmt, so spricht man von der *Maximum-Likelihood-Schätzung* (Schätzung größter Glaubwürdigkeit). Die Maximum-Likelihood-Schätzung ist eine der grundlegenden Schätzmethoden, die in der Statistik benutzt werden. Unter bestimmten Bedingungen ist diese Schätzung asymptotisch erwartungstreu, asymptotisch normalverteilt und asymptotisch effektiv. Die Maximum-Likelihood-Schätzung kann auch definiert werden, wenn die Existenz der Dichte p_Ξ nicht vorausgesetzt wird. Eine ausführliche Theorie zu diesen Fragen kann der Leser z. B. in dem Buch von WILKS [1] finden.

Da das arithmetische Mittel unabhängiger Zufallsgrößen die übliche Schätzung des Erwartungswertes in der Monte-Carlo-Methode ist, scheint es passend, einige diese Schätzung betreffende Fakten anzuführen, obwohl sie hinreichend gut bekannt sind.

Es sei ξ normalverteilt mit dem Erwartungswert a und der Dispersion σ^2; $\xi_1, \dots, \xi_N$ sind unabhängige Stichprobenwerte von ξ. Die Likelihood-Funktion in bezug auf den Parameter a ist in diesem Fall

$$p_N(x_1, \dots, x_N, a) = \frac{1}{(2\pi)^{N/2}\,\sigma^N} \prod_{i=1}^{N} \exp\left[-\frac{(x_i - a)^2}{2\sigma^2}\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - a)^2\right]. \qquad (1.3.1)$$

Ihr Maximum erreicht sie unter der Bedingung

$$Q = \sum_{i=1}^{N} (x_i - a)^2 \to \min .$$

Es ist leicht zu sehen, daß Q das Minimum genau dann erreicht, wenn $a = \hat{a}(x_1, \ldots, x_N) = N^{-1} \sum_{i=1}^{N} \xi_i$ ist. Daher ist das arithmetische Mittel (Stichprobenmittel) $\bar{\xi} = N^{-1} \sum_{i=1}^{N} \xi_i$ die Maximum-Likelihood-Schätzung für $a = \mathsf{M}\xi$. Man prüft leicht nach, daß $\bar{\xi}$ normalverteilt ist mit dem Erwartungswert a und der Dispersion σ^2/N. Unter Benutzung von Tabellenwerten für die Normalverteilung kann man ein Vertrauensintervall für a zum Konfidenzniveau γ angeben. So ist beispielsweise $\mathsf{P}(|\hat{a} - a| > 3\sigma/\sqrt{N}) = 0{,}003$.

Das arithmetische Mittel besitzt einige optimale Eigenschaften auch unter allgemeineren Voraussetzungen. Wir betrachten alle möglichen linearen erwartungstreuen Schätzungen des Erwartungswertes der Zufallsgröße ξ mit endlicher Dispersion σ^2 und im übrigen beliebiger Verteilung, also Schätzungen der Art

$$\mathscr{L} = \sum_{i=1}^{N} c_i \xi_i\,, \qquad c_i > 0\,. \tag{1.3.2}$$

Man zeigt leicht, daß in der Klasse der erwartungstreuen Schätzungen der Gestalt (1.3.2) das arithmetische Mittel die kleinste Dispersion hat. In der Tat gilt für eine erwartungstreue Schätzung notwendigerweise die Gleichung

$$\sum_{i=1}^{N} c_i = 1\,, \tag{1.3.3}$$

und die Dispersion von $\mathscr{L}$ ist gleich $\mathsf{D}\mathscr{L} = \sigma^2 \sum_{i=1}^{N} c_i^2$. Unter der Bedingung (1.3.3) hat $\mathsf{D}\mathscr{L}$ ein eindeutiges Minimum, das für $c_1 = \cdots = c_N = 1/N$ erreicht wird. Die angeführten Fakten geben die Begründung für die Schätzung des Erwartungswertes einer Zufallsgröße ξ durch das arithmetische Mittel $\bar{\xi}$ ihrer Stichprobenwerte. Als erwartungstreue Schätzung der Dispersion wird in der Statistik gewöhnlich

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\xi_i - \bar{\xi})^2 \tag{1.3.4}$$

benutzt. Hier gilt nach WILKS [1] folgendes: Sind $\xi_1, \ldots, \xi_N$ unabhängige Stichprobenwerte einer Zufallsgröße ξ, die den Erwartungswert a, die Dispersion σ^2 und auch endliche dritte und vierte Momente hat, so hat unter den Schätzungen für $\mathsf{D}\xi$ von der Gestalt $\sum_{i,j=1}^{N} c_{ij}(\xi_i - \bar{\xi})(\xi_j - \bar{\xi})$ die Schätzung

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 \tag{1.3.5}$$

die kleinste Dispersion. Eine Bereichsschätzung der Stichprobendispersion für eine normalverteilte Zufallsgröße ξ ist mit der χ^2-Verteilung verbunden. Man sagt, daß eine Zufallsgröße χ^2-verteilt mit k Freiheitsgraden ist, wenn sie die Wahrscheinlichkeitsdichte

$$p_\eta^{(k)}(x) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} \left(\frac{x}{2}\right)^{k/2-1} e^{-x/2}\,, \qquad x > 0\,, \tag{1.3.6}$$

besitzt ($\Gamma(y)$ ist die Gammafunktion). Die Summe der Quadrate von N unabhängigen normalverteilten Zufallsgrößen mit dem Erwartungswert Null und der Dispersion Eins hat eine χ^2-Verteilung mit $N - 1$ Freiheitsgraden. Daher gilt

$$\gamma = \mathsf{P}\left(a < \frac{(N-1)\,s^2}{\sigma^2} < b\right) = \int_a^b p_\eta^{(N-1)}(x)\,dx\,. \qquad (1.3.7)$$

Die Werte des Integrals auf der rechten Seite von (1.3.7) sind tabelliert. Die Gleichung (1.3.7) ist äquivalent der Gleichung

$$\mathsf{P}\left(\frac{(N-1)\,s^2}{b} < \sigma^2 < \frac{(N-1)\,s^2}{a}\right) = \gamma\,,$$

woraus folgt, daß $[(N-1)\,s^2/b,\ (N-1)\,s^2/a]$ ein Vertrauensintervall für σ^2 zum Konfidenzniveau γ ist.

In den folgenden Kapiteln werden wir eine Reihe anderer statistischer Schätzungen betrachten, die in Verbindung mit Monte-Carlo-Methoden benutzt werden.

2. Empirische Verteilungsfunktionen und Kriterien für die Übereinstimmung von Verteilungen. Es sei ξ eine Zufallsgröße mit der Verteilungsfunktion $F_\xi(x)$, und $\xi_1, \ldots, \xi_N$ seien unabhängige Realisierungen von ξ. Wir bezeichnen mit $\nu_N(x)$ die Anzahl der Werte ξ_k, für die $\xi_k < x$ ist. Die Funktion $F_N^*(x) = \nu_N(x)/N$ heißt empirische Verteilungsfunktion der Zufallsgröße ξ. Es ist bekannt (vgl. Gnedenko [1]), daß die empirische Verteilungsfunktion für $N \to \infty$ der Wahrscheinlichkeit nach in jedem Punkt x gegen $F_\xi(x)$ strebt. Uns interessiert die Verteilung des Fehlers, der beim Ersetzen von $F_\xi(x)$ durch die empirische Verteilungsfunktion entsteht. In der mathematischen Statistik gibt es eine Reihe von Kriterien, die eine Beurteilung der Annäherung von $F_N^*(x)$ an $F_\xi(x)$ erlauben.

1. *Kriterium von* A. N. Kolmogoroff. Für alle $\gamma > 0$ gilt

$$\lim_{N\to\infty} \mathsf{P}\left(\sup_x |F_\xi(x) - F_N^*(x)| < \frac{\gamma}{\sqrt{N}}\right) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\gamma^2)\,. \qquad (1.3.8)$$

2. ω^2-*Kriterium.* Für alle $\gamma > 0$ gilt

$$\lim_{N\to\infty} \mathsf{P}\left(\int [F_\xi(x) - F_N^*(x)]^2\,dF(x) < \frac{\gamma}{N}\right) = G(\gamma)\,, \qquad (1.3.9)$$

wobei $G(x)$ eine Verteilungsfunktion mit der charakteristischen Funktion

$$\mathsf{M}e^{i\lambda x} = \prod_{k=1}^{\infty}\left(1 - \frac{2i\lambda}{k^2\,\pi^2}\right)^{-1/2}$$

ist (bekanntlich bestimmt die charakteristische Funktion eindeutig die Verteilungsfunktion).

3. *Kriterium von* H. W. Smirnow. Für alle $\gamma > 0$ gilt.

$$\lim_{N\to\infty} \mathsf{P}\left(\sup_x (F_N^*(x) - F_\xi(x)) < \frac{\gamma}{\sqrt{N}}\right) = 1 - \exp(-2\gamma^2)\,. \qquad (1.3.10)$$

Die Funktionen auf den rechten Seiten der Gleichungen (1.3.8), (1.3.9) und (1.3.10) sind tabelliert.

4. χ^2-*Kriterium von* K. PEARSON. Nach diesem Kriterium besitzt die Größe

$$\chi^2 = N \sum_{j=1}^{l} \left\{ \left[\int_{\Delta_j} dF^*(x) - \int_{\Delta_j} dF(x) \right]^2 \Big/ \int_{\Delta_j} dF(x) \right\}, \tag{1.3.11}$$

wobei $\Delta_1, \Delta_2, \ldots, \Delta_l$ disjunkte Intervalle auf der reellen Achse sind, für $N \to \infty$ eine χ^2-Verteilung mit $l - 1$ Freiheitsgraden, falls die Integrale von dF und dF^* über jedes der Intervalle Δ_j nicht zu klein sind. Praktisch wird dieses Kriterium für den Fall benutzt, daß $N \int_{\Delta_j} dF(x) > 10$ ist. Werden gewisse Annahmen (sogenannte Hypothesen Δ_j) über die Gestalt der Funktion F gemacht, ergeben sich bei N gegebenen Stichprobenwerten die Prüfgrößen

$\sqrt{N} \sup_x |F(x) - F^*(x)|$ für das Kolmogoroff-Kriterium,

$N \int [F(x) - F^*(x)]^2 \, dF(x)$ für das ω^2-Kriterium usw.

Weil die Verteilung dieser Prüfgrößen bei geeigneter Wahl von F bekannt ist, kann der Apparat der statistischen Hypothesenprüfung benutzt werden, d. h., die Hypothese über die Nähe zwischen F und F^* wird angenommen, wenn die entsprechende Prüfgröße einen festgesetzten Wert nicht übersteigt, oder abgelehnt im entgegengesetzten Fall.

Die praktische Seite der Anwendung der beschriebenen Kriterien auf Aufgaben, wie sie bei der Benutzung der Monte-Carlo-Methode entstehen, wird recht ausführlich in der Monographie von GOLENKO [1] betrachtet.

3. Die Monte-Carlo-Methode. Eine grundlegende numerische Aufgabe, die gewöhnlich mit der Monte-Carlo-Methode gelöst wird, ist die Schätzung des Erwartungswertes irgendeiner Zufallsgröße, d. h. die Berechnung eines Lebesgueschen Integrals über einem Wahrscheinlichkeitsmaß. Dieses Maß kann recht komplizierter Natur sein und kann als Komposition einfacherer Verteilungen angegeben sein. Am häufigsten wurde dabei der Fall studiert, daß ein endliches zweites Moment der Zufallsgröße existiert (*Aufgabe* A, wie wir auch sagen werden). Das einfachste Rechenschema für diese Aufgabe besteht in folgendem: Man berechnet N unabhängige Realisierungen der Zufallsgröße und schätzt ihren Erwartungswert durch das arithmetische Mittel dieser Realisierungen ab. Als Grundlage dafür dient die Tatsache, daß unter den linearen erwartungstreuen Schätzungen das arithmetische Mittel die kleinste Dispersion hat. Eine Abschätzung des Fehlers mit Wahrscheinlichkeitscharakter kann man aus der Tschebyscheffschen Ungleichung (1.2.6) erhalten.

Wenn wir ein genügend kleines γ vorgeben und $\gamma = \sigma^2/N\varepsilon^2$ setzen (also $\varepsilon = \sigma/\sqrt{N\gamma}$), ergibt sich

$$\mathsf{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} \xi_i - \mathsf{M}\xi\right| < \frac{\sigma}{(N\gamma)^{1/2}}\right) \geqq 1 - \gamma\,, \tag{1.3.12}$$

d. h., mit einer Wahrscheinlichkeit von $1 - \gamma$ unterscheidet sich das arithmetische Mittel unabhängiger Realisierungen von ξ um nicht mehr von $\mathsf{M}\xi$ als $\sigma/\sqrt{N\gamma}$. Bei festem γ und σ nimmt der Fehler also wie $N^{-1/2}$ ab.

Sind die Bedingungen für die Anwendbarkeit des zentralen Grenzwertsatzes erfüllt, so kann die Größe $\frac{1}{N}\sum_{i=1}^{N}\xi_i$ für genügend großes N als normalverteilt mit dem Erwartungswert $\mathsf{M}\xi$ und der Dispersion σ^2/N angesehen werden. Das gibt uns die Möglichkeit, ein Konfidenzintervall für den Fehler zu konstruieren, wie das im ersten Abschnitt dieses Paragraphen angegeben wurde. Dafür ist aber notwendig, daß sich die Verteilung der Größe $\frac{1}{N}\sum_{i=1}^{N}\xi_i$ höchstens um eine Größe der Ordnung $N^{-1/2}$ von der Normalverteilung unterscheidet. Die Größe der Dispersion, die den Fehler bestimmt, kann auch während der Rechnung geschätzt werden. Auf diese Weise erhält man die Möglichkeit, während der Rechnung die Zahl N zu bestimmen, die die notwendige Genauigkeit mit der vorgegebenen Wahrscheinlichkeit (Sicherheit) garantiert. Diese Eigenschaft der Monte-Carlo-Methode ist also für die praktische Anwendung der Methode sehr wichtig. Die Aufgabe A ist daher eng mit der Aufgabe der Parameterschätzung einer Normalverteilung verknüpft. Hier liefert die Schätzung des Erwartungswertes die Lösung der Aufgabe, während die Schätzung der Dispersion eine Abschätzung des Fehlers gewährleistet. Da die Ordnung $O(N^{-1/2})$, mit der der Fehler abnimmt, für genauere Berechnungen zu groß ist, gibt es verschiedene Verfahren der Transformation von Zufallsgrößen, die den Erwartungswert fest lassen, aber die Dispersion ändern. Uns interessieren natürlich Verfahren zur Verkleinerung der Dispersion. Die Darstellung solcher Verfahren macht einen bedeutenden Teil dieses Buches aus.

Die Möglichkeit der genannten Transformation von Zufallsgrößen resultiert aus der Darstellung (1.2.1), also aus der Gleichung

$$\mathsf{M}\xi = \int \xi(\omega)\,\mathsf{P}(d\omega) = \int \xi(\omega)\,\frac{d\mathsf{P}}{d\mu}(\omega)\,\mu(d\omega)$$

mit $\mathsf{P} \ll \mu$. Diese Darstellung ermöglicht es, anstelle des Erwartungswertes von ξ nach dem Maß P den Erwartungswert von $\xi\frac{d\mathsf{P}}{d\mu}$ nach dem Maß μ zu berechnen.

Benutzt man $\mathsf{D}\xi = \int \xi^2(\omega)\,\mathsf{P}(d\omega) - (\mathsf{M}\xi)^2$, so gilt

$$\mathsf{D}\left(\xi\frac{d\mathsf{P}}{d\mu}\right) = \int \left(\xi\frac{d\mathsf{P}}{d\mu}\right)^2 \mu(d\omega) - (\mathsf{M}\xi)^2\,. \tag{1.3.13}$$

Daher ist im allgemeinen $\mathsf{D}\xi \neq \mathsf{D}\left(\xi\frac{d\mathsf{P}}{d\mu}\right)$. Vom numerischen Standpunkt genügt es nicht, μ so zu wählen, daß $\mathsf{D}\left(\xi\frac{d\mathsf{P}}{d\mu}\right)$ kleiner als $\mathsf{D}\xi$ wird, weil die Modellierung der Zufallsgröße $\xi\,\frac{d\mathsf{P}}{d\mu}$ bedeutend komplizierter sein kann als die Modellierung der ursprünglichen Größe ξ. Darum muß man eine Transformation benutzen, die den zum Erreichen der geforderten Genauigkeit (bei gegebener Sicherheit) notwendigen Rechenaufwand verkleinert.

Wir werden im weiteren Transformationen, die den Erwartungswert der Zufallsgröße fest lassen, aber ihre Dispersion ändern, Modifikationen in der Monte-Carlo-Methode nennen. Die Klasse von Modifikationen, die wir betrachten werden, wird nicht durch Transformationen der Gestalt (1.3.13) erschöpft. Der Arbeit von FROLOW und TSCHENZOW [1] folgend, führen wir den Begriff der effektiven Modifikation (im Rahmen des Schemas A) ein. Es sei τ die Zeit, die für die Berechnung einer Realisierung der Zufallsgröße η auf einer EDVA notwendig ist. Als Effektivität einer Modifikation werden wir die Größe $(\tau \mathsf{D}\eta)^{-1}$ einführen, d. h., der Modifikation mit kleinerem $\tau\mathsf{D}\eta$ wird eine größere Effektivität zugeschrieben. Dabei sind aber noch einige Einschränkungen zu machen. Erstens kann dieses Effektivitätskriterium natürlich nur im Rahmen des Schemas A angewendet werden. Bei der Betrachtung anderer Aufgabentypen wird darüber noch ausführlicher gesprochen. Zweitens muß gleiche Qualität der Programme vorausgesetzt werden, mit denen die Rechnung der verschiedenen Modifikationen ausgeführt wird. Somit hängt die Zeit wesentlich davon ab, ob das Programm oder der benutzte Compiler für die EDVA gewissen Qualitätsforderungen entspricht. Schließlich wird man manchmal eine weniger effektive Modifikation deswegen bevorzugen, weil es schon ausgearbeitete Algorithmen und Programme gibt.

Jetzt soll der Erwartungswert einer Zufallsgröße unter der Bedingung berechnet werden, daß ihr zweites Moment unendlich ist (*Aufgabentyp* B). Aus der Beziehung (1.3.13) folgt bereits die Möglichkeit, die Aufgabe B auf die Aufgabe A zurückzuführen, d. h. eine Zufallsgröße mit demselben Erwartungswert, aber endlicher Dispersion zu konstruieren. Manchmal ist es aber schwer, eine solche Transformation zu finden, oder es ist praktisch unmöglich, sie wegen ihrer Kompliziertheit zu realisieren. Es kann auch sein, daß der Umfang der Rechenarbeit bei vorgegebenem Fehler trotz der erreichten Endlichkeit der Dispersion wächst. (Das oben definierte Effektivitätskriterium einer Modifikation ist in diesem Fall nicht anwendbar). Die Frage der Konstruktion einer Schätzung für den Erwartungswert im Fall der Aufgabe B ist noch ungenügend untersucht. Die Benutzung des arithmetischen Mittels als Schätzwert in der bekannten Weise ist gerechtfertigt, weil die Konvergenz des Schätzungswertes gegen den Erwartungswert der Wahrscheinlichkeit nach gesichert ist (die Schätzung ist erwartungstreu und konsistent). Die Frage der Konvergenzordnung bei der Berechnung endlichdimensionaler Integrale wurde von N. S. BACHWALOW [3] untersucht. Ein späteres, aber auch allgemeineres Resultat wurde in Kap. I, § 2, in Gestalt der Ungleichung (1.2.8) formuliert. Aus dieser Ungleichung folgt

$$\mathsf{M}\left|\frac{1}{N}\sum_{i=1}^{N}\xi_i - \mathsf{M}\xi\right| \leqq \frac{2}{N^{r-1}}\mathsf{M}|\xi - \mathsf{M}\xi|^r\,, \qquad 1 \leqq r \leqq 2\,, \tag{1.3.14}$$

d. h., der Erwartungswert des Fehlerbetrages der Schätzung nimmt wie N^{1-r} ab. Die genaue Verteilung des Fehlers kann auch erhalten werden, aber ihre Form ist recht kompliziert, und in der Literatur gibt es keine fertigen Algorithmen, aus der man in bequemer Form ein Vertrauensintervall für den Fehler der Schätzung konstruieren könnte.

Falls das r-te Moment $(1 < r < 2)$ existiert, können offensichtlich bessere Schätzungen für den Erwartungswert als das arithmetische Mittel konstruiert

werden. So stellt die Benutzung des Gesetzes der großen Zahlen nur eine der Möglichkeiten bei der Anwendung der Monte-Carlo-Methode dar. Eine andere Möglichkeit bieten die verschiedenen Schätzmethoden der mathematischen Statistik. Es gibt einige Beispiele für die Lösung der Aufgabe B. Wir werden diese Beispiele in den folgenden Kapiteln untersuchen.

Das Effektivitätskriterium für Modifikationen, das in allgemeiner Form oben formuliert wurde, kann offensichtlich auf den Fall der Aufgabe B erweitert werden, aber die genaue Formulierung eines solchen Kriteriums ist wegen der noch ungenügenden Untersuchungen ein offenes Problem.

Wir werden im folgenden auch die Aufgabe betrachten, den Erwartungswert abhängiger Zufallsgrößen zu schätzen (*Aufgabe* C). Einige Fakten, die sich auf diese Aufgabe beziehen, wurden im § 2 dieses Kapitels angegeben. In der Mehrzahl der Fälle besteht das allgemeine Lösungsschema wie auch im Fall A in der Parameterschätzung einer Normalverteilung. Eine gewisse Schwierigkeit in numerischer Hinsicht bietet hier die Schätzung der Dispersion. Ausführlicher werden wir uns mit dieser Frage im nächsten Kapitel beschäftigen. Man kann auch noch eine *Aufgabe* D aufzeigen. Hier ist der Erwartungswert abhängiger Zufallsgrößen zu schätzen, wenn die Grenzverteilung des arithmetischen Mittels durch ein von der Normalverteilung abweichendes, stabiles Gesetz dargestellt wird. Dem Autor sind aber keine Beispiele der praktischen Lösung dieser Aufgabe bekannt.

Die aufgezählten Aufgaben erschöpfen natürlich nicht alle Aufgabentypen, die mit Hilfe der Monte-Carlo-Methode gelöst werden können, obwohl im gewissen Sinne andere Aufgaben auf sie zurückführbar sind. So kann z. B. die Schätzung des Erwartungswertes einer zufälligen Funktion $\xi(\vartheta, \omega)$ auf die Schätzung ihres Erwartungswertes für einige feste Parameterwerte ϑ und anschließende näherungsweise Konstruktion der Funktion $M\xi(\vartheta, \omega)$ durch Interpolation zurückgeführt werden. Hier sind auch andere Lösungswege möglich, die von uns später betrachtet werden.

Aus dem Dargelegten folgt, daß die Monte-Carlo-Methode eng mit Aufgaben der Wahrscheinlichkeitstheorie, der mathematischen Statistik und der numerischen Mathematik verknüpft ist. Im Zusammenhang mit der Modellierung von Zufallsgrößen (insbesondere gleichverteilten) spielt auch die Zahlentheorie eine wesentliche Rolle. Die mit der Zahlentheorie zusammenhängenden Fragen werden hauptsächlich im letzten Kapitel dieses Buches behandelt.

Gegenüber anderen numerischen Verfahren zeichnet sich die Monte-Carlo-Methode durch ihre Einfachheit und Allgemeinheit aus. Die langsame Konvergenz ist ein wesentlicher Nachteil dieser Methode, aber unter bestimmten Voraussetzungen kann, wie wir weiter sehen werden, durch ihre Modifikationen eine bessere Konvergenzordnung erreicht werden. Dabei wird allerdings die numerische Prozedur komplizierter und nähert sich in ihrer Kompliziertheit anderen Prozeduren der numerischen Mathematik an. Die Konvergenz der Monte-Carlo-Methode ist eine Konvergenz der Wahrscheinlichkeit nach. Dieser Umstand sollte aber kaum zu ihren Nachteilen gerechnet werden, da stochastische Methoden in genügendem Maß durch praktische Anwendungen gerechtfertigt werden. Was schließlich Aufgaben mit stochastischem Inhalt betrifft, so erscheint bei ihrer Untersuchung sogar die Konvergenz der Wahrscheinlichkeit nach im gewissen Sinn natürlich.

II. MODELLIERUNG VON VERTEILUNGEN

§ 1. Unabhängige gleichverteilte Zufallsgrößen

Wie schon bemerkt, liegt der Anwendung der Monte-Carlo-Methode eine auf dem Intervall [0, 1] gleichverteilte eindimensionale Zufallsgröße α zugrunde, d. h. eine Zufallsgröße mit der Verteilungsdichte

$$p_\alpha(x) = \begin{cases} 1 & \text{für} \quad x \in [0, 1]\,, \\ 0 & \text{für} \quad x \notin [0, 1]\,. \end{cases} \tag{2.1.1}$$

Wir geben einige einfache Eigenschaften von α an.

1. Ist α mit der Dichte (2.1.1) verteilt, so ist $\alpha' = (b - a)\,\alpha + a$ gleichverteilt auf dem Intervall $[a, b]$, $b > a$, d. h., α' besitzt die Dichte

$$p_{\alpha'}(x) = \begin{cases} \dfrac{1}{b-a} & \text{für} \quad x \in [a, b]\,, \\ \\ 0 & \text{für} \quad x \notin [a, b]\,. \end{cases}$$

2. Die diskreten Zufallsgrößen β_i $(i = 1, \ldots, N)$ seien in ihrer Gesamtheit unabhängig, die β_i $(i = 1, \ldots, N-1)$ nehmen nur die Werte 0 und 1 mit den Wahrscheinlichkeiten $\mathsf{P}(\beta_i = 0) = \mathsf{P}(\beta_i = 1) = 1/2$ an, und β_N sei auf [0, 1] gleichverteilt. Dann ist

$$\sum_{i=1}^{N-1} \frac{\beta_i}{2^i} + \frac{\beta_N}{2^{N-1}} \tag{2.1.2}$$

eine auf [0, 1] gleichverteilte Zufallsgröße.

Wir betrachten den Fall $N = 2$. Dann ergibt sich

$$\begin{aligned} \mathsf{P}\left(\frac{\beta_1}{2} + \frac{\beta_2}{2} \leqq x\right) &= \mathsf{P}(\beta_1 = 0)\,\mathsf{P}\left(\frac{\beta_2}{2} \leqq x\right) + \mathsf{P}(\beta_1 = 1)\,\mathsf{P}\left(\frac{1}{2} + \frac{\beta_2}{2} \leqq x\right) \\ &= \frac{1}{2}\,\mathsf{P}\left(\frac{\beta_2}{2} \leqq x\right) + \frac{1}{2}\,\mathsf{P}\left(\frac{\beta_2}{2} \leqq x - \frac{1}{2}\right). \end{aligned}$$

Nun ist aber

$$\mathsf{P}(\beta_2 \leqq 2x) = \begin{cases} 0 & \text{für} \quad x \leqq 0\,, \\ 2x & \text{für} \quad x \in (0, 1/2]\,, \\ 1 & \text{für} \quad x > 1/2\,, \end{cases}$$

$$\mathsf{P}(\beta_2 \leqq 2x - 1) = \begin{cases} 0 & \text{für} \quad x \leqq 1/2\,, \\ 2x - 1 & \text{für} \quad x \in (1/2, 1]\,, \\ 1 & \text{für} \quad x > 1\,. \end{cases}$$

Daraus folgt unmittelbar die gesuchte Beziehung

$$\mathsf{P}\left(\frac{\beta_1}{2}+\frac{\beta_2}{2}\leqq x\right)=\begin{cases}0 & \text{für} \quad x\leqq 0\,,\\ x & \text{für} \quad 0<x\leqq 1\,,\\ 1 & \text{für} \quad x>1\,.\end{cases}$$

Der Beweis im allgemeinen Fall wird mit vollständiger Induktion geführt.

3. Es sei $F_\alpha(x)$ die Verteilungsfunktion zu der Dichte $p_\alpha(x)$; die β_i $(i=1, \ldots, N-1)$ seien in ihrer Gesamtheit unabhängige diskrete Zufallsgrößen, die die Werte 0 und 1 mit den Wahrscheinlichkeiten $\mathsf{P}(\beta_i = 0) = \mathsf{P}(\beta_i = 1) = 1/2$ annehmen. Dann gilt

$$\mathsf{P}\left(\sum_{i=1}^{N-1}\frac{\beta_i}{2^i}\leqq x\right)-F_\alpha(x)\leqq\frac{1}{2^{N-1}}.$$

In der Tat, ist β_N auf $[0, 1]$ gleichverteilt und sind die β_i $(i = 1, \ldots, N)$ in ihrer Gesamtheit unabhängig, so gilt bei beliebigen festen Werten von β_N

$$\mathsf{P}\left(\sum_{i=1}^{N-1}\frac{\beta_i}{2^i}\leqq x\right)=\mathsf{P}\left(\frac{\alpha'-\beta_N}{2^{N-1}}\leqq x\right)=\mathsf{P}\left(\alpha'\leqq\frac{x+\beta_N}{2^{N-1}}\right),$$

wobei α' auf $[0, 1]$ gleichverteilt ist. Dann gilt aber

$$\mathsf{P}\left(\alpha'\leqq\frac{x+\beta_N}{2^{N-1}}\right)\leqq F_\alpha(x)+\frac{1}{2^{N-1}},$$

woraus das Gewünschte folgt.

Daher ist $\lim\limits_{N\to\infty}\sum\limits_{i=1}^{N-1}\frac{\beta^i}{2^i}$ auf $[0, 1]$ gleichverteilt. Man kann auch sagen, daß α die Darstellung $0{,}\beta_1\,\beta_2 \ldots$ als Dualzahl gestattet, wobei die β_i unabhängige diskrete Zufallsgrößen sind, die die Werte 0 und 1 mit gleicher Wahrscheinlichkeit annehmen. Analog gestattet α die Darstellung in der Form eines Bruches $0{,}\gamma_1\gamma_2 \ldots$ des r-Positionssystems, wobei die γ_i unabhängig sind und die Werte $0, 1, \ldots, r-1$ mit der Wahrscheinlichkeit $1/r$ annehmen.

An eines der Verfahren, eine Folge unabhängiger Realisierungen $\alpha_1, \alpha_2, \ldots$ der Zufallsgröße α mit Hilfe des Roulettespiels oder des Münzenwerfens zu gewinnen, erinnerten wir schon im ersten Kapitel. Es findet seine Begründung in den aufgeführten Eigenschaften von α. Man kann von solchen Verfahren ausgehen, um Realisierungen von α zu erhalten. Dabei entstehen aber eine Reihe von Fragen:

1. Wie ist eine ideale Durchführung der genannten Experimente zu garantieren, wenn der Einfluß unvermeidlicher Fehler bei der gewählten „Experimentiervorrichtung" ausgeschlossen werden soll?
2. Wenn man von der Gleichverteilung auf $[0, 1]$ spricht, dann hat man eine Realisierung von α mit einer unendlichen Anzahl gültiger Ziffern im Auge. Das setzt eine beliebig große Genauigkeit bei der Berechnung mit dem Roulettespiel oder eine beliebig große Anzahl von Münzenwürfen voraus, um eine Realisierung von α zu erhalten. Worin bestehen nun die Folgen der unvermeidlichen endlichen Stellenzahl?

3. Schließlich ist die Frage keineswegs trivial, wie ein analoges ideales Experiment auf einer Rechenanlage zu organiseren ist.

Die Erörterung dieser Fragen soll auch in den vorliegenden Paragraphen mit erfolgen. Vor allem werden wir uns mit der Endlichstelligkeit jeder Realisierung von α beschäftigen. Diese Frage entsteht bei der Anwendung jeder numerischen Methode und ist sogar weder mit der Undurchführbarkeit eines Experiments mit hinreichend langer Dauer oder beliebiger Genauigkeit noch mit den Möglichkeiten entsprechender Rechenanlagen verbunden. Dabei kann man von zwei Standpunkten ausgehen. Der erste besteht darin, daß wir von vornherein auf die Betrachtung einer Zufallsgröße verzichten, die auf dem Intervall $[0, 1]$ gleichverteilt ist und stattdessen eine Zufallsgröße $\tilde{\alpha}$ betrachten, die gleichverteilt in bezug auf ein geeignetes diskretes Maß ist. Wenn die Anlage eine feste Wortlänge von n Dualziffern für Zahlen aus $[0, 1]$ hat, so kann man errechnen, daß $\tilde{\alpha}$ mit der Wahrscheinlichkeit 2^{-n} die Werte $k \cdot 2^{-n}$ $(k = 0, 1, \ldots, 2^n - 1)$ annimmt, und nun die Aufgabe betrachten, Realisierungen dieser Zufallsgröße zu erhalten.

Der andere Standpunkt besteht darin, daß wir gegebenenfalls mit beliebig vielen Ziffern rechnen könnten, daß man sich aber auf n Stellen beschränkt, weil die weiteren Ziffern das Ergebnis nicht beeinflussen. Es ist rationell, diesen zweiten Standpunkt etwas hervorzuheben, denn der erste Standpunkt bedeutet eine Beschränkung analog der Forderung, bei der Berechnung von $\sqrt{2}$ nur n Ziffern zu benutzen, während das Ergebnis der Rechnungen wesentlich von den folgenden Ziffern abhängen kann. Weiterhin werden wir daher annehmen, daß α mit jeder notwendigen Anzahl von Ziffern erhalten werden kann. Dabei ist es offensichtlich, daß mit einer genügenden Anzahl von Realisierungen von $\tilde{\alpha}$ auch eine Realisierung von α mit der notwendigen Genauigkeit zu erhalten ist. Es genügt, $\alpha = \sum_{i=0}^{m} 2^{-mi}\,\tilde{\alpha}_i$ zu setzen (bei entsprechender Wahl von m und eventueller Rundung). Die Frage nach der notwendigen Anzahl von Ziffern hängt von der konkreten Aufgabe ab und muß mit Mitteln der numerischen Mathematik untersucht werden. So wird unsere Aufgabe zurückgeführt auf die Aufgabe, Realisierungen von diskreten, gleichverteilten Zufallsgrößen zu bekommen. Wenn solche Realisierungen in der notwendigen Menge erhalten werden, so kann man sie in der Rechenanlage speichern und aus ihnen dann Realisierungen mit der notwendigen Anzahl von Ziffern zusammensetzen. Anders gesagt, man kann Tabellen von Zufallszahlen benutzen.

Tabellen von Zufallszahlen wurden schon lange vor der Entstehung der Monte-Carlo-Methode in ihrer gegenwärtigen Form aufgestellt. Solche Tabellen wurden mit der Entwicklung von Verfahren der Zufallsauswahl bei der Planung von Experimenten in Biologie, Medizin, Landwirtschaft und anderen Gebieten notwendig. Es gibt verschiedene Ausgaben solcher Tabellen, die gewöhnlich in Gruppen angeordnete Zahlen des Zehnersystems enthalten. Die erste Tabelle dieser Art wurde von Tippet [1] aufgestellt, gebraucht wurden sie bei Volkszählungen. Ebensolche Tabellen wurden in der Sowjetunion von Kadyrow [1] angefertigt. Bei ihrer Aufstellung ist es sehr schwer, den Einfluß solcher Faktoren auszuschließen, die die Gleichverteilung und damit die Zufälligkeit stören. Kendall und Babington-Smith [1] benutzten für die Anfertigung ihrer Zufallszahlentabellen ein Roulette, dargestellt durch einen in zehn gleiche Sektoren geteilten

Kreis (entsprechend den zehn Ziffern 0, 1, ..., 9). Sie trafen verschiedene Vorsichtsmaßnahmen, um eine einwandfreie Durchführung des Experiments zu gewährleisten. Aber, wie die Autoren mitteilen, hegte ein bei der Arbeit helfender Assistent eine Antipathie gegenüber ungeraden Ziffern (insbesondere gegen die Ziffern 1, 3 und 9) und eine Sympathie für die Ziffern 0 und 6. Darum mußten die durch ihn registrierten Resultate ausgeschlossen werden. Durch die RAND-Corporation wurden 1947 Zufallszahlentabellen aufgestellt, die 1000000 Dezimalziffern enthalten. Dabei wurde eine spezielle schnelle Vorrichtung eines „elektronischen Roulettes" benutzt, ein Zufallszahlengenerator, der auf einem physikalischen Prinzip beruht (Generator zufälliger Impulse). Die erhaltenen Daten wurden einer sorgfältigen Analyse durch statistische Teste unterworfen. Ein gewisser Teil der Daten bestand die Prüfung nicht und erforderte eine bestimmte Korrektur des erhaltenen Materials. Wir werden hier nicht konkrete Resultate aus der Überprüfung der genannten Tabellen anführen. Diese Resultate kann man in der schon zitierten Literatur finden. Es ist wichtig zu bemerken, daß die Güte des Experiments letzten Endes auf der Grundlage statistischer Teste beurteilt werden muß.

Gewöhnlich benutzt man ein recht kompliziertes System von Testen, die sowohl auf die ganze Folge wie auch auf einzelne ihrer Teile angewendet werden.

Es sei $\beta_1, \ldots, \beta_N$ eine Folge von Realisierungen der Zufallsgröße β, die die Werte 0 und 1 annimmt. Zerlegt man diese Folge in unabhängige Gruppen der Länge n und setzt voraus, daß β annähernd gleichverteilt ist, dann kann man annehmen, daß die erhaltenen Gruppen annähernd die Dualzahlendarstellung einer gleichverteilten Zufallsgröße angeben. Die Prüfung auf Gleichverteilung und Unabhängigkeit der erhaltenen Zahlen wird mit Hilfe von Testen durchgeführt.

Insbesondere erlauben es die Kriterien der Übereinstimmung von Verteilungen (Kap. I, § 3), die Abweichung einer empirischen Verteilungsfunktion der gegebenen Folge von einer theoretischen zu beurteilen. Diese Kriterien werden auch sehr oft in der Praxis benutzt. Wir werden nicht näher auf konkrete Teste und die Technik ihrer Anwendung in bezug auf die Gleichverteilung von Folgen eingehen. Diese Dinge sind z. B. in der Monographie von Golenko [1] enthalten. Wir bemerken nur, daß die Wahl des Konfidenzniveaus bei dem genannten Verfahren schwer theoretisch zu begründen ist; und man muß solche Aussagen wie „gewöhnlich ist es üblich anzunehmen" und dergleichen verwenden.

Die Behauptung „eine gegebene Folge $\beta_1, \ldots, \beta_N$, die aus Nullen und Einsen besteht, stellt mit der Wahrscheinlichkeit 1 eine Folge unabhängiger Realisierungen der Zufallsgröße β, $\mathsf{P}(\beta = 0) = \mathsf{P}(\beta = 1) = 1/2$ dar" kann nur für $N \to \infty$ ausgesprochen werden, wenn Paare und andere Kombinationen von Nullen und Einsen in der Folge mit der Häufigkeit angetroffen werden, die gegen die entsprechende Wahrscheinlichkeit strebt. Das bedeutet, daß bei beliebigem festen n die Beziehung

$$\lim_{N \to \infty} \frac{1}{N^n} \sum_{i_1, \ldots, i_n} 1 = \mathsf{P}(\beta_1 = \gamma_1, \ldots, \beta_n = \gamma_n) \tag{2.1.3}$$

erfüllt sein muß, wobei $\beta_1, \ldots, \beta_N$ unabhängige Realisierungen von β sind, $\gamma_1, \ldots, \gamma_n$ eine gegebene Folge von Nullen und Einsen ist und i_k Werte aus einer Untermenge von $(1, \ldots, N)$ annimmt, für die $\beta_{i_k} = \gamma_k$ gilt. Diese Beziehung soll für alle $n = 1, 2, \ldots, N$ und jede mögliche Wahl der $\gamma_1, \ldots, \gamma_n$ erfüllt sein. Es ist klar, daß die

Prüfung solcher Beziehungen praktisch nicht möglich ist. Die statistischen Teste garantieren eine annähernde Erfüllung dieser Gleichungen, aber ihr Gebrauch ist mit den schon genannten Unzulänglichkeiten verbunden.

Die erste Aufgabenstellung für den Fall eines endlichen N betrachtete A. N. Kolmogoroff [2], der den Begriff von (n,ε)-Tabellen von Zufallszahlen einführte und ihre Existenz bewies. Eine Stichprobe vom Umfang n aus der endlichen (n,ε)-Tabelle vom Umfang N erfüllt mit der Genauigkeit ε die Bedingungen, die für Tabellen unendlichen Umfangs gelten. Obwohl diese Idee im Zusammenhang mit den untersuchten Fragen erstrangige Bedeutung hat, beschränken wir uns nur auf den Hinweis bezüglich der Arbeit von Kolmogoroff [2], denn eine Begründung der Monte-Carlo-Methode auf der Grundlage dieser Ideen ist praktisch noch nicht verwirklicht worden.

Entsprechend dem von uns eingenommenen Standpunkt werden wir im folgenden annehmen, daß man im Fall befriedigender Ergebnisse bei der Prüfung einer Folge mit Hilfe statistischer Teste diese Folge in Rechnungen mit der Monte-Carlo-Methode anwenden kann. Wir werden während der weiteren Darstellung aber auf diese Fragen zurückkommen und präzisieren, welche Teste gemeint sind.

Wenn es eine genügend umfangreiche Tabelle von Zufallszahlen gibt, so kann damit die Frage nach einer Quelle für unabhängige Realisierungen von α als prinzipiell gelöst betrachtet werden. Praktisch ist aber die Speicherung einer umfangreichen Tabelle im Speicher einer EDVA äußerst unbequem, so daß Tabellen von Zufallszahlen praktisch nicht benutzt werden (mit sehr seltenen Ausnahmen).

Eine bequemere Methode zur Erzeugung der Realisierungen von α besteht in der Schaffung einer speziellen Einrichtung, eines sogenannten Zufallszahlengenerators, auf physikalischer Grundlage. Ein solcher Generator gibt bei jedem Arbeitstakt eine zufällige Ziffer in den Speicher ab. Falls derartige Einrichtungen praktisch genutzt werden, besteht ein Mangel in der Unmöglichkeit, die Rechnungen zu wiederholen, und, was noch wichtiger ist, in der Störanfälligkeit der Einrichtung im Arbeitsprozeß. Daher ist eine parallele Prüfung der so erzeugten Zufallszahlenfolge durch Teste notwendig. Diese Gründe bedingen auch die seltene Anwendung physikalischer Zufallszahlengeneratoren.

In praktischen Rechnungen ist es am bequemsten, Realisierungen von α mittels irgendeines Algorithmus zu erhalten. Zahlen, die so erzeugt werden, heißen Pseudozufallszahlen. Es ist gewöhnlich recht schwierig zu zeigen, daß der eine oder andere Algorithmus eine Folge mit den von uns gewünschten Eigenschaften liefert. Daher wird ein strenger Beweis oft durch gewisse intuitive Erwägungen ersetzt oder nur einige Eigenschaften des Algorithmus werden theoretisch begründet. Weiter wendet man die schon erwähnten Teste an, auf deren Grundlage der Algorithmus entweder als brauchbar anerkannt oder verworfen wird. Dazu kommt die zusätzliche Forderung, daß für die Erzeugung der Pseudozufallszahlen nur eine kleine Anzahl von Arbeitstakten der EDVA benötigt werden sollte.

Die Mehrzahl der benutzten Pseudozufallszahlenfolgen wird durch eine Rekursionsformel auf folgende Weise erhalten. Es sei

$$x_{i+k} = a(x_{i+k-1}, \dots, x_i) \tag{2.1.4}$$

mit fester Zahl k, wobei $i = 1, 2, \dots$ ist; a stellt eine ganzzahlige Funktion ganzzahliger Argumente mit $0 \leqq a < P$ dar (P natürliche Zahl), und $x_0, x_1, \dots, x_{k-1}$ sind

natürliche Zahlen, die P nicht übersteigen und am Beginn der Rechnung gegeben sein müssen. Eine Pseudozufallszahlenfolge $\alpha_0, \alpha_1, \ldots, \alpha_i \in [0, 1]$, wird dann durch die Gleichungen

$$\alpha_i = \frac{x_i}{P}, \qquad i = 0, 1, \ldots, \tag{2.1.5}$$

definiert. In der Mehrheit der konkreten Fälle ist a in der Form

$$x_{i+k} \equiv \left(\sum_{j=0}^{k-1} b_j x_{i+j} + \vartheta \right) \pmod{P} \tag{2.1.6}$$

gegeben, wobei die b_j und ϑ ganze Zahlen sind. Als x_{i+k} nimmt man also den Rest bei der Division des Wertes einer ganzzahligen linearen Funktion durch die Zahl P. Die Methode, eine Pseudozufallsfolge durch Formel (2.1.6) zu erhalten, trägt die Bezeichnung Kongruenzmethode. Sie wurde durch Lehmer [1] vorgeschlagen. Es gibt eine umfangreiche Literatur, in der die Eigenschaften der so erhaltenen Folgen untersucht werden (siehe Jansson [1]).

Die einfachste Gestalt bekommt (2.1.6) für $k = 1$ und $\vartheta = 0$. In diesem Fall spricht man von einer multiplikativen Methode zur Erzeugung von Pseudozufallsfolgen. Für $\vartheta \neq 0$ gebraucht man die Bezeichnung „gemischte Methode".

Es ist leicht zu sehen, daß die durch die Formeln (2.1.4) und (2.1.5) erhaltenen Folgen periodisch sind, weil die Funktion a nur eine endliche Anzahl von Werten annimmt und die Kombination von Anfangswerten, die die Folge vollständig bestimmt, sich früher oder später wiederholt.

Die multiplikative Methode wurde auf Grund zahlentheoretischer Erwägungen vorgeschlagen. Zu ihrer Begründung werden wir im letzten Kapitel kommen, wo auch Ergebnisse angeführt werden, die die Periodenlänge der entsprechenden Folgen betreffen. Auf andere Methoden zur Erzeugung von Pseudozufallszahlen werden wir nicht eingehen. Angaben darüber findet der Leser beispielsweise in den Monographien von Golenko [1], Jansson [1] und in der Arbeit von Sobol [4], in denen solche Methoden recht ausführlich beschrieben werden mit Hinweisen auf Ergebnisse bei der Anwendung statistischer Teste.

Im Hinblick auf die genannten Monografien sind wir nur auf die allgemeine Problematik eingegangen, die mit der Erzeugung unabhängiger Realisierungen einer gleichverteilten Zufallsgröße verbunden ist.

Faßt man das Gesagte zusammen, so bemerkt man, daß sich natürlicherweise zwei Zugänge zur genannten Aufgabe abzeichnen. Einen von ihnen kann man etwa den „experimentellen Weg" nennen. Er besteht darin, daß irgendein Experiment durchgeführt wird, dessen Charakter beliebig ist und von dem anzunehmen ist, daß es Werte der geforderten Folge liefert. Danach gestatten statistische Teste, die Berechtigung unserer Annahme zu beurteilen.

Das System der Teste muß genügend umfangreich sein, wenn wir unsere Folge tatsächlich als Folge unabhängiger Realisierungen der Zufallsgröße α betrachten wollen. Es ist übrigens zu bemerken, daß manchmal zur Lösung der gegebenen Aufgabe oder sogar einer gewissen Klasse von Aufgaben die Prüfung der Folge mit einigen speziellen Testen genügt.

Der zweite mögliche Weg, den wir im Kern nicht berührt haben und der natürlicherweise konstruktiv zu nennen ist, besteht in der Angabe eines theoretischen

Verfahrens zur Konstruktion einer Folge mit den notwendigen Eigenschaften. Diese Aufgabe ist sehr kompliziert, und wir betrachten sie im letzten Kapitel.

Im weiteren nehmen wir an, daß uns eine Folge von Realisierungen einer auf $[0, 1]$ gleichverteilten Zufallsgröße zur Verfügung steht, und bei der Herleitung theoretischer Ergebnisse (bis zum siebenten Kapitel) gehen wir von dieser Annahme aus. Durch den Buchstaben α mit verschiedenen Indizes werden die genannten Realisierungen bezeichnet.

Bei der Beschreibung von Algorithmen ist es aber angenehm, ein gewisses konkretes Verfahren zur Erzeugung einer Folge von Realisierungen $\alpha_1, \alpha_2, \ldots$ im Auge zu haben. Dazu dient eine Rekursionsformel (2.1.5), (2.1.6), in der die Parameter so gewählt wurden, daß die Periode der erhaltenen Folge genügend groß ist und die ausgewählten statistischen Teste ein befriedigendes Resultat liefern. Dabei ist folgendes zu bemerken. Wenn die Pseudozufallszahlen auf einer Maschine erhalten wurden, die im Dualzahlsystem mit n Dualziffern β_i arbeitet (gewöhnlich ist dann $P = 2^n$), so kann sich die Verteilung der letzten Ziffern wesentlich von der diskreten $\mathsf{P}(\beta_i = 1) = \mathsf{P}(\beta_i = 0) = 1/2$ unterscheiden. Mehr noch, die letzte Ziffer dieser Zahlen ist immer gleich 1. Die statistischen Teste, die auf Pseudozufallszahlen wie auf Realisierungen einer auf $[0, 1]$ gleichverteilten Zufallsgröße anwendbar sind, können nichtsdestoweniger befriedigende Ergebnisse liefern, weil der Beitrag der letzten Ziffern entsprechend gering ist. Das muß man bei der Benutzung einzelner Ziffern von Pseudozufallszahlen sowie bei der Erzeugung von Pseudozufallszahlen durch Verschiebung der Ziffern beachten.

Ist $\boldsymbol{\alpha} = (\alpha^{(1)}, \ldots, \alpha^{(s)})$ eine im s-dimensionalen Einheitswürfel gleichverteilte Zufallsgröße, so kann man für ihre Realisierung eine beliebige unabhängige Auswahl von s unabhängigen Realisierungen α_k der auf $[0, 1]$ gleichverteilten Zufallsgröße α nehmen. Analog erhält man Realisierungen der Zufallsgröße $\boldsymbol{\alpha}'$, die im s-dimensionalen Parallelepiped $\{a_i \leqq x_i \leqq b_i\}_{i=1,\ldots,s}$ gleichverteilt ist, indem man beispielsweise

$$\boldsymbol{\alpha}'_k = \big(a_1 + (b_1 - a_1)\,\alpha_k, \ldots, a_s + a_s + (b_s - a_s)\,\alpha_{k+s-1}\big)$$

setzt. Werden dabei für die Prüfung der Qualität der Folge Übereinstimmungskriterien benutzt, so ist mindestens notwendig, ihre entsprechenden s-dimensionalen Verallgemeinerungen zu gebrauchen, um die Nähe zwischen empirischer und theoretischer Verteilung zu garantieren (aber nicht die Unabhängigkeit einzelner Realisierungen!).

Will man unabhängige Realisierungen eines im s-dimensionalen Gebiet $\mathscr{D}$ des euklidischen Raumes gleichverteilten Zufallsvektors erhalten, also eines Zufallsvektors mit der Verteilungsdichte

$$p_\alpha(x_1, \ldots, x_s) = \begin{cases} 1/|\mathscr{D}| & \text{für} \quad (x_1, \ldots, x_s) \in \mathscr{D}\,, \\ 0 & \text{sonst} \end{cases}$$

($|\mathscr{D}|$ bezeichnet den Inhalt des Gebietes $\mathscr{D}$), so geht man gewöhnlich folgendermaßen vor. Ist $\mathscr{D}$ beschränkt, so kann man $\mathscr{D}$ in ein s-dimensionales Parallelepiped $\{a_i \leqq x_i \leqq b_i\}_{i=1,\ldots,n}$ einbetten. Man nimmt nun eine Folge von Realisierungen von $\boldsymbol{\alpha}'$ und entfernt jene Glieder daraus, die nicht in $\mathscr{D}$ liegen. Ausführlicher werden wir auf diese Frage im nächsten Paragraphen eingehen.

§ 2. Methoden zur Modellierung allgemeiner Verteilungen

Wir nehmen an, daß es ein Verfahren zur Erzeugung unabhängiger Realisierungen einer auf [0, 1] gleichverteilten Zufallsgröße gibt. Dann betrachten wir davon ausgehend allgemeine Methoden zur Erzeugung unabhängiger Realisierungen von Zufallsgrößen mit gegebenem Verteilungsgesetz. Die theoretischen Grundlagen dieser Methoden sind recht einfach. Wir werden sie in Form von Sätzen mit Beweis oder unter Hinweis auf die entsprechende Literatur formulieren. Außerdem werden wir Algorithmen betrachten, die für viele Fälle gewisse Bedeutung haben.

Eine Zufallsgröße ξ sei durch ihre Verteilungsfunktion $F_\xi(x)$ und entsprechend ein Zufallsvektor $\Xi = (\xi_1, \ldots, \xi_s)$ durch die Verteilungsfunktion $F_\Xi(x_1, \ldots, x_s)$ gegeben. Dann gilt der

Satz 1 (Transformation von N. W. Smirnow). *Genügt ξ der Gleichung*

$$\int_{-\infty}^{\xi} dF_\xi(t) = \alpha\,, \qquad \xi = F_\xi^{-1}(\alpha)\,, \tag{2.2.1}$$

wobei α eine auf [0, 1] *gleichverteilte Zufallsgröße ist, so ist ξ nach dem Gesetz $F_\xi(x)$ verteilt.*

In der Tat gilt

$$\mathsf{P}(\xi \leqq x) = \mathsf{P}\left(\int_{-\infty}^{\xi} dF_\xi(t) \leqq \int_{-\infty}^{x} dF_\xi(t)\right),$$

denn $F_\xi(t)$ ist eine nichtfallende nichtnegative Funktion. Weil ξ Lösung der Gleichung (2.2.1) ist, gilt weiter $\mathsf{P}(\xi \leqq x) = \mathsf{P}(\alpha \leqq F_\xi(x))$. Da aber α auf [0, 1] gleichverteilt und $0 \leqq F_\xi(x) \leqq 1$ ist, erhält man schließlich $F_\xi(x) = \mathsf{P}(\alpha \leqq F_\xi(x))$.

Für einen Zufallsvektor gilt ein analoges Resultat.

Satz 2. *Die Zufallsgrößen $\xi_1, \ldots, \xi_s$ besitzen die gemeinsame Verteilungsfunktion $F_\Xi(x_1, \ldots, x_s)$, und es gelten die Beziehungen*

$$F_1(x_1) = \mathsf{P}(\xi_1 \leqq x_1) = \underbrace{\int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{s} dF_\Xi(x_1, \ldots, x_s)\,,$$

$$F_2(x_1, x_2) = \mathsf{P}(\xi_2 \leqq x_2 \mid x_1)$$
$$= \underbrace{\int_{-\infty}^{x_2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{s-1} dF_\Xi(x_1, \ldots, x_s) \left[\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{s-1} dF_\Xi(x_1, \ldots, x_s)\right]^{-1},$$

. . . .

$$F_s(x_1, \ldots, x_s) = \mathsf{P}(\xi_s \leqq x_s \mid x_1, \ldots, x_{s-1})$$
$$= \int_{-\infty}^{x_s} dF_\Xi(x_1, \ldots, x_s) \left[\int_{-\infty}^{\infty} dF_\Xi(x_1, \ldots, x_s)\right]^{-1}.$$

Dann sind die Größen $\xi_1, \ldots, \xi_s$ Lösungen der Folge von Gleichungen

$$F_1(\xi_1) = \alpha_1 , \qquad F_2(\xi_1, \xi_2) = \alpha_2, \ldots, \qquad F_s(\xi_1, \ldots, \xi_s) = \alpha_s \tag{2.2.2}$$

und besitzen $F_\Xi(x_1, \ldots, x_s)$ als gemeinsame Verteilungsfunktion.

Die Gleichungen (2.2.2) heißen auch Umkehrformeln.

Der Beweis verläuft analog zum Beweis im eindimensionalen Fall.

Algorithmus[1]). Die Gleichungen (2.2.1) und (2.2.2) haben immer eine eindeutige Lösung, weil F_ξ bzw. F_Ξ eine nichtfallende Funktion ist. Ist F_ξ unstetig, so muß man zur Berechnung einer bestimmten Lösung die Tatsache benutzen, daß F_ξ rechtsseitig stetig ist. Praktischerweise stellt man die Lebesgue-Stieltjes-Integrale in den Formeln (2.2.1) und (2.2.2) in Gestalt zweier Summanden dar, als Lebesguesches Integral und eine dem diskreten Anteil entsprechende Summe.

Ausführlicher gehen wir auf zwei Fälle der Modellierung einer eindimensionalen Zufallsgröße mit Hilfe der Formel (2.2.1) ein:

a) Die Dichte $p_\xi(x) = \dfrac{dF_\xi(x)}{dx}$ existiert. Dann hat (2.2.1) die Gestalt

$$\int_{-\infty}^{\xi} p_\xi(x)\, dx = \alpha . \tag{2.2.3}$$

Ist eine analytische Berechnung des Integrals möglich, so ergibt sich für die Bestimmung von ξ im allgemeinen eine transzendente Gleichung. Ohne Kenntnis konkreter Eigenschaften der Funktion p_ξ ist es schwer, ein effektives Lösungsverfahren vorzuschlagen. Die Benutzung des Newtonschen Verfahrens beispielsweise kann bei kleinem p_ξ eine zu langsame Konvergenz aufweisen. Noch komplizierter wird die Aufgabe, wenn das Integral mit einer Quadraturformel näherungsweise berechnet werden muß. Daher kann man sagen, daß Formel (2.2.1) nur in hinreichend einfachen Fällen ein effektives Ergebnis liefert. Zum Beispiel erhalten wir für die Exponentialverteilung

$$p_\xi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für} \quad x \geqq 0 , \\ 0 & \text{für} \quad x < 0 , \end{cases} \qquad \lambda > 0 , \tag{2.2.4}$$

aus (2.2.3) $e^{-\lambda\xi} = 1 - \alpha$. Indem man die Größe $1 - \alpha$, die auch auf $[0, 1]$ gleichverteilt ist, durch α ersetzt, erhält man

$$\xi = -\frac{1}{\lambda} \log \alpha . \tag{2.2.5}$$

---

[1]) Wir setzen der Einfachheit halber voraus, daß sich F_ξ bzw. F_Ξ nur aus einem stetigen und einem diskreten Anteil zusammensetzen. Dem Autor ist der Fall unbekannt, daß bei der Lösung von konkreten Aufgaben die Notwendigkeit entsteht, Verteilungen mit singulärem Anteil zu modellieren. Es ist auch zu bemerken, daß die Unbestimmtheit (auf einer Menge vom Maß Null) bei der Bestimmung der bedingten Wahrscheinlichkeit praktisch keine Schwierigkeiten hervorruft. Der Algorithmus muß so eingerichtet sein, daß Ereignisse, die der genannten Nullmenge entsprechen, nicht eintreten.

Es muß aber bemerkt werden, daß bei der Inkaufnahme eines zusätzlichen Arbeitsaufwandes die Formeln (2.2.1), (2.2.2) eine Quelle sehr effektiver Algorithmen für die Modellierung sein können. Ein großer Rechenaufwand ergibt sich durch die Notwendigkeit, die Gleichungen (2.2.1) bzw. (2.2.2) während der Rechnung mehrfach für verschiedene Werte α aus $[0, 1]$ zu lösen. Ist F_ξ eine hinreichend glatte Funktion, so kann man diese Gleichung zunächst für eine gewisse konkrete Auswahl von α-Werten lösen und eine Tabelle von Werten der Umkehrfunktion $F_\xi^{-1}(\alpha)$ aufstellen. Werte für beliebige α können danach durch Interpolation gefunden werden. Analoge Überlegungen sind verständlicherweise auch im mehrdimensionalen Fall gültig. Aber mit dem Wachsen der Dimensionszahl wächst auch der Umfang der Tabellen stark an. Allgemein führt eine Erhöhung der Effektivität bei der Modellierung in der Regel entweder zu einem größeren Speicherbedarf der EDVA oder zur Komplizierung der Programmstruktur und zu einem umfangreicheren Programm. Eine Ausnahme stellt der Fall dar, daß $F_\xi^{-1}(\alpha)$ durch genügend einfache Funktionen gegeben ist, beispielsweise durch Polynome nicht zu hohen Grades oder durch rationale Funktionen. Besonders effektiv sind solche Prozeduren im mehrdimensionalen Fall, obwohl es nicht allzuviele konkrete Anwendungsbeispiele gibt. Derartige Beispiele werden später behandelt.

b) Es sei ξ eine diskrete Zufallsgröße, die die Werte τ_j mit den Wahrscheinlichkeiten p_j $(j = 1, 2, \ldots)$ annimmt. Dann folgt aus (2.2.1), daß $\xi = \tau_{j_0}$ ist, wobei j_0 durch die Ungleichungen

$$\sum_{j=1}^{j_0-1} p_j < \alpha \leqq \sum_{j=1}^{j_0} p_j \qquad \text{oder} \qquad 0 < \alpha - \sum_{j=1}^{j_0-1} p_j \leqq p_{j_0} \tag{2.2.6}$$

bestimmt ist. Die Ungleichungen (2.2.6) haben eine einfache geometrische Bedeutung. Das Intervall $[0, 1]$ wird in Intervalle der Länge p_j $(\sum p_j = 1)$ in der Reihenfolge der Numerierung der p_j zerlegt. Nun wird $j = j_0$ gewählt, wenn α in das entsprechende Intervall fällt. Der Algorithmus erfordert in diesem Fall keine weitere Erläuterung.

In numerischer Hinsicht einfacher ist meistens eine Methode, die VON NEUMANN [1] vorgeschlagen hat. Sie wird durch folgende Aussage begründet.

Satz 3. *Es sei $\varphi(\boldsymbol{x})$ mit $\boldsymbol{x} = (x_1, \ldots, x_s)$ eine nichtnegative Funktion, die auf dem s-dimensionalen Parallelepiped*

$$\mathcal{D} = \{a_i \leqq x_i \leqq b_i\}_{i=1,\ldots,s}$$

definiert und dort beschränkt, d.h. $\varphi(\boldsymbol{x}) \leqq M$, und integrierbar ist. Sind $\boldsymbol{\alpha}' = (\alpha_1', \ldots, \alpha_s')$ und α unabhängige, auf $\mathcal{D}$ bzw. $[0, 1]$ gleichverteilte Zufallsgrößen, so fällt die Dichte der bedingten Verteilung

$$\mathsf{P}(\alpha_i' \leqq x_i;\ i = 1, \ldots, s \mid \varphi(\boldsymbol{\alpha}') \geqq M\alpha)$$

mit $c\varphi(\boldsymbol{x})$ zusammen, wobei c eine Normierungskonstante ist.

Beweis.

$$\mathsf{P}(x_i < \alpha_i' \leqq x_i + \Delta x_i;\ i = 1, \ldots, s \mid \varphi(\boldsymbol{\alpha}') \geqq M\alpha)$$
$$= c \int_{x_1}^{x_1+\Delta x_1} dt_1 \cdots \int_{x_s}^{x_s+\Delta x_s} dt_s \int_0^1 dt\, \chi(t_1, \ldots, t_s\colon \varphi(t_1, \ldots, t_s) \geqq Mt)$$
$$= c \int_{x_1}^{x_1+\Delta x_1} dt_1 \cdots \int_{x_s}^{x_s+\Delta x_s} dt_s\, \varphi(t_1, \ldots, t_s)\,.$$

Hierbei bedeutet $\chi(t_1, \ldots, t_s\colon \varphi(t_1, \ldots, t_s) \geqq Mt)$ die charakteristische Funktion der Menge von Veränderlichen $t_1, \ldots, t_s$, für die die Ungleichung $\varphi(t_1, \ldots, t_s) \geqq Mt$ erfüllt ist. Analog findet man

$$c^{-1} = \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_s \int_0^1 \chi(t_1, \ldots, t_s : \varphi(t_1, \ldots, t_s) \geqq Mt)\, dt$$
$$= \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_s\, \varphi(t_1, \ldots, t_s)\,.$$

Daraus folgt dann die Behauptung.

Algorithmus. Ein Rechenverfahren liegt auf der Hand. Es sei $\alpha_1, \alpha_2, \ldots$ die ursprüngliche Folge unabhängiger Realisierungen von α. Wir wählen (beispielsweise) nacheinander Gruppen von $s + 1$ Zahlen:

$$\alpha_{k(s+1)+1}, \ldots, \alpha_{(k+1)(s+1)}\,, \qquad k = 0, 1, \ldots, \tag{2.2.7}$$

aus. Wir verwerfen solche Gruppen, für die die Ungleichung

$$\varphi(a_1 + (b_1 - a_1)\,\alpha_{k(s+1)+1}, \ldots, a_s + (b_s - a_s)\,\alpha_{(k+1)(s+1)}) \geqq M\alpha_{(k+1)(s+1)} \tag{2.2.8}$$

nicht erfüllt ist. Sind $k_1, k_2, \ldots$ die Werte k, für die (2.2.8) erfüllt ist, so erhält man

$$\xi_i^{(r)} = a_r + (b_r - a_r)\,\alpha_{k_i(s+1)+r}\,, \qquad r = 1, 2, \ldots, s\,,$$

als Komponenten unabhängiger Realisierungen der Zufallsgröße Ξ, die mit der Wahrscheinlichkeitsdichte $c\,\varphi(\boldsymbol{x})$ verteilt ist.

Es sei N die Anzahl von Gruppen (2.2.7), für die die Ungleichung (2.2.8) geprüft wurde, und es erwies sich dabei, daß N_1 Gruppen diese Ungleichung erfüllen. Die Größe $\lim\limits_{N\to\infty} (N_1/N)$, die offensichtlich gleich $\left[cM \prod\limits_{i=1}^{s} (b_i - a_i)\right]^{-1}$ ist, heißt Effektivität der v.-Neumannschen Methode. Sie entspricht dem Verhältnis des im Mittel nützlichen Rechenaufwandes zum gesamten Aufwand bei einer Realisierung der Zufallsgröße Ξ. Verfahren zur Erhöhung der Effektivität der v.-Neumannschen Methode werden unten untersucht. Es gilt auch die folgende Umkehrung von Satz 3 (vgl. Michailow [1]), die in einigen Fällen nützlich sein kann. Die Bezeichnungen entsprechen denen von Satz 3.

Satz 4. *Ist Ξ eine mit der Dichte $c\varphi(\boldsymbol{x})$ verteilte Zufallsgröße und ξ für jedes $\Xi = \boldsymbol{x}$ auf $[0, c\varphi(\boldsymbol{x})]$ gleichverteilt, so ist der Zufallsvektor (Ξ, ξ) in dem durch die Ungleichungen*

$$a_1 \leqq x_1 \leqq b_1, \ldots, a_s \leqq x_s \leqq b_s\,, \qquad 0 \leqq x_{s+1} \leqq c\varphi(x_1, \ldots, x_s)\,,$$

bestimmten $(s+1)$-dimensionalen Gebiet gleichverteilt.

Gilt $\varphi(\boldsymbol{x}) \geqq \varepsilon > 0$ in $\mathcal{D}$, so verläuft der Beweis ähnlich wie der Beweis von Satz 3 und kann elementar geführt werden. Wird aber $\varphi(x)$ Null, so muß man die Menge aller der Punkte von $\mathcal{D}$, in denen das eintritt, ausschließen. Der Satz kann nützlich sein bei der Modellierung einer Gleichverteilung in einem komplizierten Gebiet. Ein entsprechender Algorithmus ist offensichtlich.

In vielen Fällen erweisen sich auch verschiedene Verallgemeinerungen der v.-Neumannschen Methode als nützlich (Romanowski [2], Butler [1]). Speziell gilt der

Satz 5. *Es sei Ξ ein Zufallsvektor mit der Verteilungsfunktion $F(x_1, \ldots, x_s)$, η eine eindimensionale Zufallsgröße mit der stetigen Verteilungsfunktion $G(y)$, deren Werte dem Intervall $[0, M]$ angehören, und $r(x_1, \ldots, x_s)$ eine stetige nichtnegative Funktion mit $\sup_x r(\boldsymbol{x}) = M$. Dann wird die bedingte Verteilungsfunktion von Ξ unter der Bedingung $\eta \leqq r(\Xi)$ durch die Gleichung*

$$\begin{aligned} H(x_1, \ldots, x_s) &= \mathsf{P}(\xi_1 \leqq x_1, \ldots, \xi_s \leqq x_s \mid r(\Xi) \geqq \eta) \\ &= c \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_s} G(r(\boldsymbol{x}))\, dF(x) \end{aligned} \tag{2.2.9}$$

bestimmt, wobei c eine Normierungskonstante ist, die durch

$$c = \frac{1}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G(r(\boldsymbol{x}))\, dF(x)}.$$

erklärt ist.

Beweis. Es ist

$$\begin{aligned} &\mathsf{P}(\xi_1 \leqq x_1, \ldots, \xi_s \leqq x_s \mid r(\Xi) \geqq \eta) \\ &= c \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_s} dF(\boldsymbol{x}) \int_{-\infty}^{\infty} \chi\left(\boldsymbol{x} : r(\boldsymbol{x}) \geqq y\right) dG(y) = c \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_s} dF(\boldsymbol{x}) \int_{0}^{r(\boldsymbol{x})} dG(y)\,. \end{aligned}$$

Daraus folgt unmittelbar die Behauptung, wenn man berücksichtigt, daß c gleich

$$\frac{1}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dF(\boldsymbol{x}) \int_{-\infty}^{\infty} \chi(\boldsymbol{x} : r(\boldsymbol{x}) \geqq y)\, dG(y)}$$

gesetzt wird.

Der Algorithmus entspricht mit einigen offensichtlichen Veränderungen dem Algorithmus der v.-Neumannschen Methode. Ausführlicher betrachten wir einige Spezialfälle von Satz 5 und die entsprechenden Algorithmen.

a) Es sei $\Xi = \xi$ eine eindimensionale Zufallsgröße mit diskreter Verteilung, die die m Werte τ_i mit den gleichen Wahrscheinlichkeiten $1/m$ annimmt. Es ist nun eine Zufallsgröße zu modellieren, die dieselben Werte mit den Wahrscheinlichkeiten p_i annimmt. Wir wählen als $r(x)$ eine Funktion, die für $x = \tau_i$ durch die Gleichung $r(\tau_i) = p_i$ definiert ist, und als η eine auf $\left[0, \max\limits_i p_i\right]$ gleichverteilte Zufallsgröße. Weil in diesem Fall $G(x) = x/\max\limits_i p_i$ und $c = m \cdot \max\limits_i p_i$ ist, folgt aus (2.2.9) die Gleichung $dH = mr(x)\, dF(x)$.

Algorithmus. Die Modellierung einer Zufallsgröße, die mit gleichen Wahrscheinlichkeiten m gegebene Werte τ_i annimmt, führt auf die Realisierung einer ganzen Zahl i_0 und des ihr zugeordneten τ_{i_0}. In diesem Fall wird i_0 auf folgende Weise bestimmt. Die Realisierung α wird mit m multipliziert und der ganze Teil der erhaltenen Zahl gleich i_0 gesetzt. Ist i_0 bestimmt, so prüft man die Ungleichung $p_{i_0} \geqq \alpha \max\limits_j p_j$, und im Fall ihrer Erfüllung ist die Realisierung von ξ gleich τ_{i_0}. Andernfalls wird ein neuer α-Wert bestimmt und die Rechnung wiederholt.

Die erhaltene Prozedur ist das diskrete Analogon der v.-Neumannschen Methode. Ihre Effektivität, die wie im stetigen Fall definiert wird, ist gleich $1/m \cdot \max\limits_i p_i$.

b) $F_\Xi(\boldsymbol{x})$ besitzt die Dichte $\varphi(\boldsymbol{x})$. Dann hat auch H eine Dichte $h(\boldsymbol{x})$, die durch die Gleichung

$$h(\boldsymbol{x}) = c\varphi(\boldsymbol{x})\, G\big(r(\boldsymbol{x})\big) \tag{2.2.10}$$

bestimmt ist. Im eindimensionalen Fall, wenn $r(x) = x$ und auch η Werte aus $[0, 1]$ annimmt, ergibt sich

$$h(x) = c\varphi(x)\, G(x)\,. \tag{2.2.11}$$

Nehmen wir an, daß $a = 1$ und η auf $[0, 1]$ gleichverteilt ist, aber r eine beliebige, den Bedingungen von Satz 5 genügende Funktion ist, so gilt

$$h(x) = \frac{\varphi(x)\, r(x)}{\int \varphi(x)\, r(x)\, dx} \tag{2.2.12}$$

unter der Bedingung $\int \varphi(x)\, r(x)\, dx < \infty$.

In diesem Fall ist es leicht, eine Funktion $r(x)$ zu ermitteln, damit die Gleichung (2.2.12) eine gegebene Funktion h bestimmt. Diese Funktion $r(x)$ wird bis auf einen konstanten Faktor durch die Gleichung

$$r(x) = \begin{cases} c_1 h(x)/\varphi(x) & \text{für} \quad h(x) > 0 \quad \text{und} \quad \varphi(x) > 0\,, \\ 0 & \text{sonst} \end{cases}$$

bestimmt. Ist $\max\limits_x \big(h(x)/\varphi(x)\big) = 1$, so ist $c_1 = 1$. Sonst setzen wir voraus, daß $h(x)$ beschränkt ist und $\varphi(x)$ der Ungleichung $0 < \varepsilon \leqq \varphi(x)$ für alle x im betrachteten Gebiet genügt. Dann muß bei der Wahl von c_1 berücksichtigt werden, daß $\eta \leqq 1$ ist und folglich $r(x)$ den Wert 1 nicht übersteigen soll. Offensichtlich genügt

es, $c_1 = \min_x (\varphi(x)/h(x))$ zu setzen. Die Effektivität der Methode wird durch das Verhältnis

$$\frac{1}{\int \varphi(x)\, r(x)\, dx}$$

bestimmt.

Es lassen sich leicht weitere Spezialfälle angeben, bei denen die Formel (2.2.9) benutzt wird.

Es ist offensichtlich, daß die v.-Neumannsche Methode mit der Umkehrformel kombiniert werden kann. Die Lösung der einen oder anderen Gleichung (bei der Umkehrformel) kann ersetzt werden durch die Modellierung der entsprechenden eindimensionalen Verteilungen nach der v.-Neumannschen Methode. Die Umkehrformel ihrerseits kann bei der Benutzung von Verallgemeinerungen der v.-Neumannschen Methode vorteilhaft sein.

Die folgende Methode, die die Bezeichnung „Methode der festen Ordinaten" trägt (ROMANOWSKI [2]) wird durch nachstehenden Satz begründet.

Satz 6. *Es sei $r(x)$ eine nichtnegative, auf dem Intervall $[a, b]$ definierte und dort stetige Funktion mit $\max_x r(x) = 1$. Wählen wir dann eine Realisierung α' der auf $[0, 1]$ gleichverteilten Zufallsgröße α und außerdem von α' und in der Gesamtheit unabhängige Realisierungen $\alpha_1', \alpha_2', \ldots$ der auf $[a, b]$ gleichverteilten Zufallsgröße, bis das erste Mal die Ungleichung*

$$r(\alpha_i') > \alpha' \tag{2.2.13}$$

erfüllt ist, so hat α_i' die Verteilungsdichte[1])

$$p(x) = \int_0^{r(x)} \frac{b-a}{\operatorname{mes}\{x : r(x) \geqq y\}}\, dy\,. \tag{2.2.14}$$

Beweis. Es sei $r(\alpha_i') \geqq \alpha$. Dann gilt

$$\mathsf{P}(x \leqq \alpha_i' \leqq x + \Delta x \mid \alpha = y) = \begin{cases} \Delta x/m(y) & \text{für} \quad r(x) \geqq y\,, \\ 0 & \text{sonst}\,, \end{cases}$$

wobei $m(y) = (b-a)^{-1} \operatorname{mes}\{x : r(x) \geqq y\}$ ist. Daraus ergibt sich

$$\mathsf{P}(x \leqq \alpha_i' \leqq x + \Delta x \mid r(\alpha_i') \geqq \alpha') = \int_x^{x+\Delta x} dx \int_0^{r(x)} \frac{b-a}{\operatorname{mes}\{x : r(x) \geqq y\}}\, dy\,.$$

Algorithmus. Man wählt zunächst irgendeine Realisierung von α und danach Realisierungen $\alpha_1', \alpha_2', \ldots$ und prüft die Ungleichung $\alpha' < r(\alpha_k')$. Ist k_0 der erste k-Wert, für den die genannte Ungleichung erfüllt ist, so wird α_{k_0}' eine Realisierung der mit der Dichte $p(x)$ verteilten Zufallsgröße.

[1]) Mit mes wird wie üblich das gewöhnliche Maß einer Menge bezeichnet.

Spezialfall. Ist $p(x)$ monoton, so muß auch $r(x)$ monoton sein. Beide seien nun fallend. Dann gilt

$$p(x) = (b - a) \int_0^{r(x)} [r^{-1}(y)]^{-1} \, dy \, . \tag{2.2.15}$$

Für differenzierbare $p(x)$ und $r(x)$ erhalten wir

$$p'(x) = (b - a) \, r'(x) \, [r^{-1}(r(x))]^{-1} \, .$$

Wegen $r^{-1}(r(x)) = x - a$ ergibt sich

$$p'(x) = \frac{b - a}{x - a} \, r'(x) \qquad \text{und} \qquad r(x) = \int_a^x \frac{x - a}{b - a} \, p'(x) \, dx + c \, .$$

Da $r(x)$ monoton fällt und $\max\limits_x r(x) = 1$ ist, können wir $r(a) = 1$, d. h. $c = 1$, setzen und finden nach partieller Integration

$$r(x) = \frac{1}{b - a} \, [(x - a) \, p(x) - F(x)] + 1 \, , \tag{2.2.16}$$

wobei $F(x)$ die zur Dichte $p(x)$ gehörende Verteilungsfunktion ist. In Einzelfällen kann die Formel (2.2.16) als Quelle bequemer Algorithmen dienen, obwohl im allgemeinen die notwendige Benutzung von $F(x)$ ihre Anwendung praktisch unbequem macht. Man kann leicht die Klasse von Verteilungsfunktionen $F(x)$ angeben, für die $r(x)$ eine genügend einfache Gestalt hat.

Es sei dazu $a = 0$, $b = 1$ und $g(x)$ eine differenzierbare Funktion, deren Werte verhältnismäßig einfach zu berechnen sind. Die Funktion $r(x)$ ist leicht zu berechnen, wenn $xp(x) - F(x) = r(x) - 1 = g(x)$ gesetzt wird. $F(x)$ muß also der Differentialgleichung $xF'(x) - F(x) = g(x)$ genügen, deren formale Lösung $F(x) = x \int t^{-2} g(t) \, dt + C$ ist. Die Modellierung von Verteilungen solcher Art ist nicht trivial.

Im allgemeineren Fall, also ohne die Monotonievoraussetzung über $p(x)$, wird die Funktion $h^{-1}(y) = \text{mes} \, \{x \colon p(x) > y\}$ eingeführt. Die zugehörige Umkehrfunktion $h(y)$ kann als Dichte einer gewissen Zufallsgröße betrachtet werden. Da $h(x)$ monoton ist, kann mit ihrer Hilfe unter Benutzung von Formel (2.2.16) eine Funktion $\varrho(x)$ konstruiert werden. Wegen $\varrho^{-1}(y) = \text{mes} \, \{x : r(x) \geqq y\}$ ist $r(x) = \varrho(h^{-1}(p(x)))$. Eine weitere Verallgemeinerung des beschriebenen Verfahrens besteht in der Einführung einer beliebigen Abszissenverteilung. Ist die Abszisse mit der Dichte $f_1(x)$ verteilt, so erhalten wir für monotone Funktionen die Formel (2.2.15) in der Gestalt

$$p(x) = \int_0^{r(x)} \left[\int_a^{r^{-1}(y)} f_1(t) \, dt \right]^{-1} dy \tag{2.2.17}$$

und die Formel (2.2.16)

$$r(x) = p(x) \, F_1(x) + 1 - \int_0^x p(x) \, f_1(x) \, dx \, , \tag{2.2.18}$$

wobei $F_1(x)$ eine Verteilungsfunktion mit der Dichte $f_1(x)$ ist.

Wir beschäftigen uns nun mit der sogenannten *Kompositionsmethode*, ohne den entsprechenden einfachen Satz zu formulieren. Diese von Butler [1] vorgeschlagene Methode ist ein Verfahren zur Erhöhung der Effektivität bei der Modellierung von Verteilungen durch Einführung von mehrstufigen Modellierungen und durch Zerlegung der zu modellierenden Verteilung in einfachere Verteilungen. Wir führen hier die eindimensionale Variante an. Die Verallgemeinerung auf den mehrdimensionalen Fall bereitet keine Schwierigkeiten.

Es sei η eine Zufallsgröße mit der Verteilungsfunktion $G(y)$ und $\psi(x, y)$ eine Familie von Verteilungen, die für jedes x bezüglich einer beliebigen Dichte y normiert sind. Wir nehmen an, daß die uns interessierende Dichte $\varphi(x)$ in der Gestalt

$$\varphi(x) = \int \psi(x, y)\, dG(y)$$

darstellbar ist. Wählen wir nun eine Realisierung η_0 der Zufallsgröße η und modellieren damit die Dichte $\psi(x, \eta_0)$, so erhalten wir eine Realisierung ξ, die mit der Dichte $\varphi(x)$ verteilt ist. Butler benutzte diese Methode zur Modellierung einer Dichte, die durch eine Reihe mit nichtnegativen Koeffizienten gegeben ist:

$$\varphi(x) = \sum_{i=0}^{\infty} a_i x^i\,, \qquad a_i \geqq 0\,, \quad 0 \leqq x \leqq 1\,. \tag{2.2.19}$$

Da $\varphi(x)$ normiert ist, gilt

$$\sum_{i=0}^{\infty} \frac{a_i}{i+1} = 1\,.$$

Wählen wir als η eine diskrete Zufallsgröße, die die Werte $i = 0, 1, \ldots$ mit den Wahrscheinlichkeiten $a_i/(i+1)$ annimmt, und setzen $\psi(x, y) = (y+1)\, x^y$, $0 \leqq x \leqq 1$, so gilt offenbar

$$\int \psi(x, y)\, dG(y) = \sum_{i=0}^{\infty} a_i x^i\,.$$

Die Prozedur der Modellierung besteht daher in der Bestimmung einer Realisierung i der Zufallsgröße η und danach in der Modellierung der Dichte $(i+1)\, x^i$.

Oft wird die Kompositionsmethode in folgender Weise angewendet. $\varphi(x)$ wird als Summe von zwei Summanden dargestellt:

$$\varphi(x) = p_1\varphi_1(x) + p_2\varphi_2(x)\,,$$

wobei $\varphi_1(x)$ und $\varphi_2(x)$ normierte Verteilungen sind und $p_1 > 0$, $p_2 > 0$, $p_1 + p_2 = 1$ gilt. Ist $\varphi_1(x)$ eine einfach zu modellierende Dichte, erfordert $\varphi_2(x)$ bei der Modellierung einen größeren Aufwand und gilt $p_1 \gg p_2$, so spart man beträchtliche Rechenarbeit, weil der Algorithmus häufig die Bildung von $\varphi_1(x)$ (mit der Wahrscheinlichkeit p_1) und selten die von $\varphi_2(x)$ erfordert. Diese Überlegungen werden auch zur Erhöhung der Effektivität bei der v.-Neumannschen Methode genutzt. Ist in diesem Fall die Effektivität bei der Modellierung der Dichte $\varphi(x)$ klein, so stellt man sie in der Form $\sum_{i=1}^{n} p_i\varphi_i(x)$ mit positiven p_i, $\sum_{i=1}^{n} p_i = 1$, dar, wobei die $\varphi_i(x)$ solche Dichten sind, die entweder nach der v.-Neumannschen Methode effektiv zu modellieren sind oder nur mit kleinem Faktor p_i vorkommen. Eine

Reihe von Einzelheiten, die die Erhöhung der Effektivität der v.-Neumannschen Methode betreffen, kann der Leser in der Arbeit von GOLENKO [1] finden. Im Zusammenhang mit der Kompositionsmethode erwähnen wir noch eine Arbeit von BUTLER [1], in der die Dichte $\varphi(x)$ in der Form $\varphi(x) = \sum_{i=1}^{n} a_i \varphi_i(x)\, \psi_i(x)$ mit normierten Dichten $\varphi_i(x)$ und $\varphi_i(x)\, \psi_i(x)$ dargestellt wird. Der Index i der Dichte $\varphi_i(x)\, \psi_i(x)$ wird, wie oben beschrieben, gefunden, und das Produkt $\varphi_i(x)\, \psi_i(x)$ wird mit Hilfe der verallgemeinerten v.-Neumannschen Methode (2.2.9) modelliert.

§ 3. Spezielle Methoden zur Modellierung allgemeiner Verteilungen

Die Einteilung der Modellierungsmethoden in allgemeine und spezielle ist bekanntlich relativ. Die Smirnowtransformation (Umkehrformel), die Methode der konstanten Ordinaten und andere Methoden, die in § 2 untersucht wurden, sind zweckmäßigerweise unter Beachtung der Besonderheiten der Verteilungen anzuwenden, und die gelungene Anwendung dieser oder jener allgemeinen Methode auf eine gegebene Verteilung kann als spezielle Methode betrachtet werden. Zu den allgemeinen Methoden rechnen wir solche, die es gestatten, beliebige Verteilungen oder eine umfangreiche Klasse von Verteilungen theoretisch zu modellieren. Spezielle Methoden sind solche, die auf eine recht enge Klasse oder auf spezielle Verteilungen anwendbar sind. Verständlicherweise hat es nur Sinn, solche konkreten Verteilungen zu betrachten, die oft in realen Aufgaben angetroffen werden, z. B. die Poisson- und die Normalverteilung und mit ihnen zusammenhängende Verteilungen. Bei der Modellierung von Verteilungen spezieller Art werden in breitem Maße von der Wahrscheinlichkeitstheorie her bekannte Beziehungen zwischen den Verteilungen von Zufallsgrößen ausgenutzt. Aber auch die Grenzwertsätze und einige Formeln der numerischen Analysis werden benutzt.

Insbesondere werden wir im weiteren die folgenden bekannten Beziehungen anwenden (vgl. PROCHOROW und ROSANOW [1]).

1. Ist die Zufallsgröße η eine Funktion der Zufallsgrößen $\xi_1, \ldots, \xi_s$, also $\eta = f(\xi_1, \ldots, \xi_s)$, und $\varphi(x_1, \ldots, x_s)$ die gemeinsame Verteilungsdichte der $\xi_1, \ldots, \xi_s$, so gilt

$$\mathsf{P}(c \leqq \eta \leqq d) = \int_{c \leqq f(x_1, \ldots, x_s) \leqq d} \cdots \int \varphi(x_1, \ldots, x_s)\, dx_1 \cdots dx_s\,. \tag{2.3.1}$$

2. Haben die Zufallsgrößen $\xi_1, \ldots, \xi_s$ und $\eta_1, \ldots, \eta_s$ die gemeinsamen Verteilungsdichten $p_\Xi = (x_1, \ldots, x_s)$ bzw. $p_H(y_1, \ldots, y_s)$ und ist $\eta_i = f_i(\xi_1, \ldots, \xi_s)$, $i = 1, \ldots, s$, eine Transformation mit von Null verschiedener Funktionaldeterminante

$$\Phi = \frac{\partial(f_1, \ldots, f_s)}{\partial(x_1, \ldots, x_s)}\,,$$

so sind die Dichten p_Ξ und p_H durch die Beziehung

$$p_H(y_1, \ldots, y_s) = p_\Xi\big(f_1(x_1, \ldots, x_s), \ldots, f_s(x_1, \ldots, x_s)\big)\, |\Phi| \tag{2.3.2}$$

verknüpft.

3. Sind ξ_1 und ξ_2 unabhängige Zufallsgrößen mit den Dichten $p_1(x)$ bzw. $p_2(x)$, so ist die Verteilungsdichte $p(y)$ der Summe $\xi_1 + \xi_2$ gleich (Faltungsformel)

$$p(y) = \int_{-\infty}^{\infty} p_1(y - x)\, p_2(x)\, dx\,. \tag{2.3.3}$$

Diese und einige andere Beziehungen erlauben es oft, für die Realisierung einer Zufallsgröße ξ einen unmittelbar einleuchtenden Ausdruck der Form

$$\xi = f(\alpha_1, \ldots, \alpha_n) \tag{2.3.4}$$

zu erhalten. Eine solche Darstellung der Realisierungen von ξ ist in numerischer Hinsicht sehr bequem. Wir werden zunächst besonders einfache Beispiele für die Darstellung von ξ durch eine Formel der Gestalt (2.3.4) anführen.

Beispiel 1. Aus (2.3.3) folgt unmittelbar, daß $\xi = \alpha_1 - \alpha_2$ die Verteilungsdichte

$$p_\xi(x) = \begin{cases} 1 - |x| & \text{für} \quad x \in [-1, +1]\,, \\ 0 & \text{sonst} \end{cases}$$

besitzt.

Beispiel 2. Für $\xi^{(n)} = \max(\alpha_1, \ldots, \alpha_n)$ erhält man

$$p_{\xi^{(n)}}(x) = \begin{cases} nx^{n-1} & \text{für} \quad x \in [0, 1]\,, \\ 0 & \text{sonst.} \end{cases}$$

In der Tat gilt für $\xi^{(2)} = \max(\alpha_1, \alpha_2)$

$$\begin{aligned} \mathsf{P}(\xi_2 \leqq x) &= \mathsf{P}(\alpha_1 \leqq x \mid \alpha_2 > \alpha_1) + \mathsf{P}(\alpha_2 \leqq x \mid \alpha_2 \leqq \alpha_1) \\ &= \int_0^x dx \int_0^1 \chi(x : x > y)\, dy + \int_0^x dy \int_0^1 \chi(y : y \geqq x)\, dx = \int_0^x 2x\, dx\,. \end{aligned}$$

Analoge Untersuchungen können für beliebiges n und für den Fall $\tilde{\xi}^{(n)} = \min(\alpha_1, \ldots, \alpha_n)$ durchgeführt werden. Für $\tilde{\xi}^{(n)} = \min(\alpha_1, \ldots, \alpha_n)$ ergibt sich

$$p_{\tilde{\xi}^{(n)}}(x) = \begin{cases} n(1 - x)^{n-1} & \text{für} \quad x \in [0, 1]\,, \\ 0 & \text{sonst}\,. \end{cases}$$

Von G. A. Michailow [1] wurden einzelne Klassen von Zufallsgrößen untersucht, für deren Modellierung man solche bequemen Formeln der Gestalt (2.3.4) benutzen kann. Wir führen hier einige seiner Resultate an.

Satz 7. *Die Verteilungsdichte $\varphi(x)$ sei stetig und monoton fallend für $x > 0$ und gleich Null für $x < 0$ und besitze den nicht notwendig endlichen Grenzwert*

$$\lim_{x \to +0} \varphi(x) = \mathsf{M}\,.$$

Ist die Zufallsgröße η mit der Wahrscheinlichkeitsdichte $\psi(y) = \varphi^{-1}(y)$, $0 < y < \mathsf{M}$, verteilt und α auf $[0, 1]$ gleichverteilt, so ist die Zufallsgröße

$$\xi = \alpha \varphi^{-1}(\eta) \tag{2.3.5}$$

mit der Dichte $\varphi(x)$ verteilt.

Beweis. Es ist klar, daß $\psi(y)$ auch eine Dichtefunktion ist. (Die Vertauschung der Koordinatenachsen ändert nicht den zwischen den Achsen und der Kurve $y = \varphi(x)$ eingeschlossenen Flächeninhalt.) Bei festem $\eta = y_0$ ist ξ auf dem Intervall $[0, \varphi^{-1}(y_0)]$ gleichverteilt. Dann ist aber der zufällige Vektor (ξ, η) im Gebiet

$$(0 \leqq y \leqq M; 0 \leqq y \leqq \varphi^{-1}(y))$$

und (vgl. Satz 4) folglich auch im Gebiet

$$(0 < x < \infty; 0 < y < \varphi(x))$$

gleichverteilt. Nach dem erwähnten Satz 4 bedeutet das aber, daß ξ mit der Dichte $\varphi(x)$ verteilt ist.

Beispiele.

1. $\varphi(x) = -\log x\,, \quad 0 < x \leqq 1\,, \quad \xi = \alpha_1 \alpha_2\,;$

2. $\varphi(x) = \arccos x\,, \quad 0 \leqq x \leqq 1\,, \quad \xi = \alpha_1 \sqrt{1 - \alpha_2^2}\,.$

Analogen Charakter trägt der

Satz 8 (Michailow [2]). *Die Dichte $\varphi(x)$ sei stetig differenzierbar, gleich Null für $x \leqq 0$, und es sei $\varphi(x) + \varphi'(x) > 0$ für $x > 0$. Dann ist*

$$\psi(y) = \varphi(y) + \varphi'(y) = [\varphi(y)\, e^y]'\, e^{-y}$$

eine Dichtefunktion. Ist η eine mit der Dichte $\psi(y)$ verteilte Zufallsgröße, so ist $\xi = -\log \alpha + \eta$ mit der Dichte $\varphi(x)$ verteilt.

Beweis.

1. Es ist

$$\int_0^\infty \psi(y)\, dy = \lim_{A \to \infty} \int_0^A [\varphi(y) + \varphi'(y)]\, dy$$
$$= \lim_{A \to \infty} \left[\int_0^A \varphi(y)\, dy + \varphi(A) \right] = 1 + \lim_{A \to \infty} \varphi(A)\,.$$

Da $\psi(y) \geqq 0$ ist, existiert das uneigentliche Integral, oder es ist bestimmt divergent. Weil aber $\varphi(y)$ eine Dichtefunktion ist, kommt nur $\varphi(A) \to 0$ für $A \to \infty$ in Frage. Dann hat das Integral über ψ den Wert 1, und ψ ist auch eine Dichtefunktion.

2. Die Verteilungsdichte von $-\log \alpha$ ist e^{-z} $(z > 0)$. Nach der Faltungsformel (2.3.3) findet man für die Dichte $p_\xi(t)$ der Zufallsgröße ξ

$$p_\xi(t) = \int_0^t [\varphi(y)\, e^y]'\, e^{-y}\, e^{-t+y}\, dy = e^{-t} \int_0^t [\varphi(y)\, e^y]'\, dy = \varphi(t)\,,$$

was zu beweisen war.

Beispiel.

$$\varphi(x) = ce^{-x} \sinh\left(2\sqrt{x}\right),$$

$$\psi(y) = c[e^{-y} \sinh \sqrt{y}\, e^{y}]'\, e^{-y} = c\, e^{-y}\, y^{-1/2} \cosh \sqrt{y}\,.$$

Eine solche Verteilungsdichte hat die Zufallsgröße $\eta = (u + \sqrt{2})^2/2$, wobei u normalverteilt mit dem Erwartungswert Null und der Dispersion Eins ist. Dann ist also

$$\xi = \left[\frac{(u + \sqrt{2})^2}{2} - \log \alpha\right]. \tag{2.3.6}$$

Eine prinzipielle Möglichkeit, die Darstellung von Zufallsgrößen in der Form (2.3.4) zu untersuchen, gibt Formel (2.3.1). Aber ihre unmittelbare Benutzung führt gewöhnlich zu sehr komplizierten Berechnungen. Daher können den Sätzen 7 und 8 analoge Sätze in konkreten Fällen bedeutendes Interesse beanspruchen. Als nächstes betrachten wir nun Methoden zur Modellierung gewisser konkreter Verteilungen, die oft in der Praxis vorkommen.

1. Die Exponentialverteilung. Die entsprechende Dichte ist

$$p_\xi(x) = \begin{cases} f(x) \exp\left(-\int\limits_0^x f(t)\, dt\right) & \text{für} \quad x \geqq 0\,, \\ 0 & \text{für} \quad x < 0\,, \end{cases} \tag{2.3.7}$$

wobei $f(x)$ eine nichtnegative, über jedem endlichem Intervall $[0, C]$ integrierbare Funktion mit

$$\lim_{C\to\infty} \int\limits_0^C f(t)\, dt = +\infty$$

ist.

Mit der Exponentialverteilung eng verknüpft sind die *Poissonverteilung* und der *Poissonsche Zufallsprozeß*. Die Poissonverteilung ist eine diskrete Verteilung. Die entsprechende Zufallsgröße nimmt die Werte $i = 0, 1, \ldots$ mit der Wahrscheinlichkeit $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$ an. Der Poissonsche Prozeß ist ein Markoffscher Prozeß mit folgender Dichte der Übergangswahrscheinlichkeit:

$$p(x, y) = \begin{cases} f(y - x) \exp\left(-\int\limits_0^{y-x} f(t)\, dt\right) & \text{für} \quad y \geqq x \\ 0 & \text{für} \quad y < x\,. \end{cases} \tag{2.3.8}$$

Die Differenz $y - x$ ist also exponentialverteilt. $f(x)$ heißt Parameter des Poissonschen Prozesses. Ist die Anfangsverteilung auch eine Exponentialverteilung mit dem Parameter $f(x)$, so bildet eine Folge von Summen $\sum\limits_{i=1}^{n} \xi_i$, $n = 1, 2, \ldots$, mit unabhängigen exponentialverteilten ξ_i eine Realisierung des Poissonschen Pro-

zesses mit dem Parameter $f(x)$. Ist $f(x) \equiv \lambda$ konstant, dann bereitet die Modellierung der Dichte (2.3.7) keine Schwierigkeiten (vgl. (2.2.5)). Ist aber $f(x)$ eine verhältnismäßig komplizierte Funktion, so führt die Anwendung einer beliebigen allgemeinen Methode auf mühevolle numerische Prozeduren.

In diesem Fall ist die folgende Methode nützlich (COLEMAN [1]). Es sei $f(x) \leqq A = \text{const}$ und ξ eine mit der Dichte

$$p_\xi(x) = A \exp(-A\,x)\,, \quad x \geqq 0\,,$$

verteilte Zufallsgröße. $\xi_1, \xi_2, \ldots$ seien unabhängige Realisierungen von ξ. Man bildet dann $\zeta_i = \sum_{j=1}^{i} \xi_j$, $i = 1, 2, \ldots$ Mit einer Folge $\alpha_1, \alpha_2, \ldots$ wird die Erfüllung der Ungleichung $A\,\alpha_i \leqq f(\zeta_i)$ geprüft. Ist i_0 der erste i-Wert, für den diese Ungleichung erfüllt ist, so stellt ζ_{i_0} eine Realisierung der mit der Dichte (2.3.8) verteilten Zufallsgröße dar. Tatsächlich wird die Verteilungsfunktion von ζ_{i_0} durch die Gleichung

$$\mathsf{P}(\zeta_{i_0} < x) = \sum_{s=1}^{\infty} \mathsf{P}(E_s)$$

bestimmt, wobei die E_s Ereignisse sind, die in der Erfüllung einer Reihe von Ungleichungen

$$E_s = \{\alpha_1 > \sigma(\zeta_1), \alpha_2 > \sigma(\zeta_2), \ldots, \alpha_{s-1} > \sigma(\zeta_{s-1}), \alpha_s \leqq \sigma(\zeta_s), \zeta_s \leqq x\}$$

mit $\sigma(x) = f(x)/A$ bestehen.

Bei festem ξ_i ($\xi_1 = x_1, \ldots, \xi_s = x_s$) und folglich festem ζ_i, $i = 1, \ldots, s$, gilt wegen der Unabhängigkeit der $\alpha_1, \ldots, \alpha_s$

$$\begin{aligned} &\mathsf{P}\{E_s \mid \xi_i = x_i \ (i = 1, \ldots, s)\} \\ &\quad = \prod_{k=1}^{s-1} \mathsf{P}\left\{\alpha_k > \sigma\left(\sum_{i=1}^{k} \xi_i\right) \middle| \xi_i = x_i \ (i = 1, \ldots, k)\right\} \\ &\quad\quad \times \mathsf{P}\left\{\alpha_s \leqq \sigma\left(\sum_{i=1}^{s} \xi_i\right) \middle| \xi_i = x_i \ (i = 1, \ldots, s)\right\} \\ &\quad = \sigma\left(\sum_{i=1}^{s} x_i\right) \prod_{k=1}^{s-1} \left[1 - \sigma\left(\sum_{i=1}^{k} x_i\right)\right]. \end{aligned}$$

Weil die ξ_i unabhängig und identisch mit der Dichte $A \exp(-Ax_i)$ verteilt sind, gilt

$$\begin{aligned} \mathsf{P}(E_s) = \int_0^x dx_1 \int_0^{x-x_1} dx_2 \cdots \int_0^{x - \sum_{i=1}^{s-1} x_i} dx_s\, g(x_1)\, g(x_1 + x_2) \cdots \\ \cdots g\left(\sum_{k=1}^{s-1} x_i\right) \sigma\left(\sum_{i=1}^{s} x_i\right) A^s \exp\left(-A \sum_{i=1}^{s} x_i\right) \end{aligned} \quad (2.3.9)$$

mit $g(x) = 1 - \sigma(x)$.

Man setzt nun $y_j = \sum_{i=1}^{j} x_i$, $j = 1, \ldots, s$. Weil die x_i sich in den Grenzen $0 \leqq x_i \leqq x - \sum_{k=1}^{i-1} x_k$ ändern, findet man für die y_j die Ungleichung $y_{j-1} \leqq y_j \leqq x$. Durch Änderung der Integrationsreihenfolge in (2.3.9) ergibt sich dann

$$\mathsf{P}(E_s) = A^s \int_0^x \sigma(y_s)\, e^{-Ay_s} \int_0^{y_s} g(y_{s-1})\, dy_{s-1} \cdots \int_0^{y_2} g(y_1)\, dy_1 \,.$$

Durch sukzessives Differenzieren verifiziert man leicht, daß

$$\int_0^{y_s} g(y_{s-1})\, dy_{s-1} \cdots \int_0^{y_2} g(y_1)\, dy_1 = \frac{1}{(s-1)!}\left[\int_0^{y_s} g(y)\, dy\right]^{s-1}$$

gilt. Daraus folgt schließlich

$$\mathsf{P}(\zeta_{i_0} \leqq x) = \int_0^x \left\{ f(y)\, e^{-Ay} \sum_{s=1}^{\infty} \frac{A^{s-1}}{(s-1)!} \left[\int_0^y g(y)\, dy\right]^{s-1} \right\} dy$$

$$= \int_0^x f(y)\, e^{-Ay} \exp\left\{\int_0^y A[1-\sigma(y)]\, dy\right\} dy$$

$$= \int_0^x f(y) \exp\left\{-\int_0^y f(t)\, dt\right\} dy \,.$$

2. Die eindimensionale Normalverteilung. Es existiert eine umfangreiche Bibliographie, die sich mit Fragen zur Modellierung der Normalverteilung und damit zusammenhängender Verteilungen beschäftigt. Auf die grundlegenden Methoden gehen wir anschließend ein. Eine ausführlichere Übersicht findet man in den Monographien von GOLENKO [1], JANSSON [1], BUSLENKO u. a. [1].

Eine Reihe von Methoden benutzt den zentralen Grenzwertsatz. Wir betrachten die Summe $\sum_{i=1}^{n} \alpha_i$. Ihr Erwartungswert ist gleich $n/2$, ihre Dispersion $n/12$. Folglich hat die Zufallsgröße

$$\xi^{(n)} = \sqrt{\frac{12}{n}} \sum_{i=1}^{n} \left(\alpha_i - \frac{1}{2}\right) \tag{2.3.10}$$

den Erwartungswert Null und die Dispersion Eins und ihre Verteilung strebt für $n \to \infty$ gegen die Normalverteilung. Praktisch kann man annehmen, daß für $n = 12$ eine genügend gute Annäherung an die Normalverteilung erreicht wird. Dann hat (2.3.10) die besonders einfache Form

$$\xi^{(12)} = \sum_{i=1}^{12} \alpha_i - 6 \,. \tag{2.3.11}$$

Freilich muß man bei der Lösung konkreter Aufgaben berücksichtigen, daß für verhältnismäßig große Abweichungen die Größe $\xi^{(12)}$ keine guten Resultate liefern kann. (Für $x > 6$ verschwindet die Dichte $p(x)$ der Zufallsgröße $\xi^{(12)}$ identisch.) Ein gutes Beispiel für eine Aufgabe, in der dieser Umstand sehr wesentlich ist, findet sich in der Arbeit von Kutai und Fainschtein [1]. In einer Reihe von Fällen wird $\xi^{(n)}$ mit einer noch kleineren Anzahl von Summanden benutzt, aber man benutzt dann spezielle Transformationen, die genauere Ergebnisse liefern. Transformationen solcher Art hat Bolschew [1] untersucht. Eine der einfachsten wird durch die Formel

$$\eta^{(n)} = \xi^{(n)} + \frac{1}{20n}\left((\xi^{(n)})^3 - 3\xi^{(n)}\right) \tag{2.3.12}$$

gegeben, wobei $\xi^{(n)}$ in (2.3.10) bestimmt wird. In dem Buch von Golenko [1] wird behauptet, daß die Verteilung von $\eta^{(n)}$ bereits für $n = 5$ der Normalverteilung genügend nahe kommt.

Die Formeln (2.3.10) bis (2.3.12) haben die Gestalt (2.3.4). Zu Formeln dieser Art gehören auch Formeln, die anscheinend Wiener als erster betrachtet hat (siehe auch Jansson [1], Kronmal [1]). Sie werden nicht durch den zentralen Grenzwertsatz, sondern durch die Beziehung (2.3.2) begründet und haben die Form

$$\xi = \sqrt{-2\log\alpha_1}\cos(2\pi\alpha_2)\,, \qquad \eta = \sqrt{-2\log\alpha_1}\sin(2\pi\alpha_2)\,. \tag{2.3.13}$$

Die Zufallsgrößen ξ und η sind unabhängig und normalverteilt.

Setzt man nämlich $x = r\cos\vartheta$, $y = r\sin\vartheta$ und ist $\varphi(x, y)$ die gemeinsame Verteilungsdichte von x und y, so sind nach (2.3.2) r und ϑ mit der Dichte $r\varphi(r\cos\vartheta, r\sin\vartheta)$ verteilt. Ist daher

$$\varphi(x, y) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x^2 + y^2)\right\},$$

so sind r mit der Dichte $re^{-r^2/2}$, $r \geqq 0$ verteilt und ϑ auf $[0, 2\pi]$ gleichverteilt. Daraus folgt unter Beachtung von (2.2.1) unsere Behauptung. Freilich ist die mit der Dichte $re^{-r^2/2}$ verteilte Zufallsgröße mit Hilfe der Umkehrformel nicht einfach zu modellieren, indem man sie in der Form $\sqrt{-2\log\alpha_1}$ darstellt. Die Berechnung der Quadratwurzel und des Logarithmus sind recht zeitraubende Operationen. Man könnte dazu verschiedene Modifikationen der v.-Neumannschen Methode benutzen. Analog sind $\cos(2\pi\alpha_2)$ und $\sin(2\pi\alpha_2)$ Projektionen des Einheitsvektors, dessen Richtung in der Ebene gleichverteilt ist. In Abschnitt 5 werden wir auf eine spezielle Methode zur Berechnung der Projektionen solcher Vektoren hinweisen.

3. Die mehrdimensionale nichtausgeartete Normalverteilung von Zufallsgrößen mit dem Erwartungswert Null wird durch die Kovarianzmatrix $C = \left((c_{ij})\right)_{i,j=1,\dots,s}$, $\det C \neq 0$, bestimmt. Ihre Dichte ist

$$\varphi(x_1, \dots, x_s) = \frac{(\det C)^{-1/2}}{(2\pi)^{s/2}}\exp\left\{-\frac{1}{2}\boldsymbol{x}^\top C^{-1}\boldsymbol{x}\right\}, \tag{2.3.14}$$

wobei $\boldsymbol{x}^\mathsf{T} = (x_1, \ldots, x_s)$ ein Zeilenvektor und $\boldsymbol{x}$ der entsprechende Spaltenvektor sind. Nach (2.3.2) überführt eine nichtsinguläre lineare Transformation der Zufallsgrößen die Dichte (2.3.14) in eine Dichte derselben Art mit anderer Kovarianzmatrix. Weil die Matrix C^{-1} symmetrisch und positiv definit ist, kann sie nach dem Verfahren von CHOLESKY (FADDEJEW und FADDEJEWA [1], S. 181) in der Form SS^T mit einer Dreiecksmatrix S dargestellt werden (S^T ist die zu S transponierte Matrix). Dann ergibt sich

$$C^{-1} = (SS^\mathsf{T})^{-1} = (S^\mathsf{T})^{-1}S^{-1} \qquad \text{und} \qquad \boldsymbol{x}^\mathsf{T}C^{-1}\boldsymbol{x} = (S^{-1}\boldsymbol{x})^\mathsf{T}\, S^{-1}\boldsymbol{x} \,.$$

Wir setzen nun $\boldsymbol{y} = S^{-1}\boldsymbol{x}$, was der linearen Transformation der mit der Dichte (2.3.14) verteilten Zufallsgrößen $\xi^{(1)}, \ldots, \xi^{(s)}$ entspricht. Dann gilt offenbar

$$\varphi(x_1, \ldots, x_s) = \tilde{\varphi}(y_1, \ldots, y_s) = (2\pi)^{-s/2} \exp\{-\tfrac{1}{2}\, \boldsymbol{y}^\mathsf{T}\boldsymbol{y}\} \,. \tag{2.3.15}$$

Daraus folgt, daß die mit der Dichte $\tilde{\varphi}$ verteilten Zufallsgrößen $\eta^{(1)}, \ldots, \eta^{(s)}$ unabhängig und normalverteilt mit den Erwartungswerten Null und den Dispersionen Eins sind. Für die Erzeugung einer Realisierung der mehrdimensionalen Zufallsgrößen $(\xi^{(1)}, \ldots, \xi^{(s)})$ muß man also unabhängige Realisierungen von s eindimensionalen normalverteilten Zufallsgrößen $\eta^{(1)}, \ldots, \eta^{(s)}$ verwenden und danach die Beziehung $\Xi = SH$ ausnutzen, wobei Ξ und H Vektoren mit den Komponenten $\xi^{(i)}$ bzw. $\eta^{(i)}$ sind.

Die Dreiecksgestalt der Matrix S erlaubt es, dabei eine beträchtliche Anzahl von Speicherplätzen der EDVA zu sparen, wenn s groß ist.

4. Mit der Normalverteilung verknüpfte Verteilungen. Für eine Reihe von Zufallsgrößen, die unmittelbar durch eine normalverteilte Zufallsgröße ausgedrückt werden können, kann auf natürliche Weise ein Algorithmus für die Modellierung angegeben werden. Beispiele dafür sind Zufallsgrößen, die eine Γ-, β- oder χ^2-Verteilung besitzen (unter bestimmten Einschränkungen für die auftretenden Parameter). Die Beziehungen zwischen den genannten Verteilungen und der Normalverteilung werden in vielen Handbüchern der mathematischen Statistik angeführt (z. B. WILKS [1] u. a.). Wir beschränken uns auf die Betrachtung zweier konkreter Beispiele. Ausführlicher wird diese Frage in der Arbeit von ROMANOWSKI [2] untersucht.

a) Die χ^2-Verteilung wurde schon in (1.3.6) besprochen. Sind die ξ_i unabhängige, mit den Parametern (0, 1) normalverteilte Zufallsgrößen, so heißt die Größe $\chi_s^2 = \sum_{i=1}^{s} \xi_i^2$ mit s Freiheitsgraden χ^2-verteilt. Die Methode ihrer Modellierung erfordert keine besonderen Erläuterungen. Wird für die Modellierung der ξ_i die Formel (2.3.13) benutzt, so erhält man für gerades s

$$\chi_s^2 = -\log(\alpha_1\alpha_2 \cdots \alpha_{s/2}) \tag{2.3.16}$$

und für ungerades s

$$\chi_s^2 = -\log(\alpha_1\alpha_2 \cdots \alpha_{(s-1)/2}) - \log \alpha_{(s+1)/2} \cos^2(2\pi\alpha_{(s+3)/2}) \,.$$

b) Die Zufallsgröße $\beta_{r,s}$, die eine β-Verteilung mit den Parametern r und s (r und s natürliche Zahlen) besitzt, ist durch die Beziehung

$$\beta_{r,s} = \frac{\chi^2_{2(r+s)}}{\chi^2_{2r} + \chi^2_{2s}}$$

mit den χ^2-Verteilungen verknüpft. Daraus ergibt sich mit (2.3.16) eine Formel zur Erzeugung von Realisierungen der Größe $\beta_{r,s}$:

$$\beta_{r,s} = \frac{\log(\alpha_1\alpha_2 \cdots \alpha_{r+s})}{\log(\alpha_{r+s+1} \cdots \alpha_{2(r+s)})}. \tag{2.3.17}$$

Die Verteilungsdichte der $\beta_{r,s}$ ist

$$p_{\beta_{r,s}}(x) = \begin{cases} \dfrac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\, x^{r-1}(1-x)^{s-1} & \text{für} \quad x \in [0, 1]\,, \\ 0 & \text{sonst.} \end{cases} \tag{2.3.18}$$

5. Modellierung gleichverteilter Richtungen im Raum. Die Kugel $\sum\limits_{i=1}^{s} (x^{(i)})^2 = 1/4$ mit dem Radius 1/2 sei eingebettet in den Einheitswürfel $\{-1/2 \leqq x^{(i)} \leqq 1/2\}_{i=1,\ldots,s}$. Nun betrachten wir eine Realisierung der Größe $\boldsymbol{\alpha}'$ ($\alpha_i' = \alpha_i - 1/2$), die in dem Würfel gleichverteilt ist. Zeigt es sich, daß

$$\sum_{i=1}^{s} (\alpha_i')^2 \geqq 1/4 \tag{2.3.19}$$

ist, so wählt man neue, von den vorherigen unabhängigen α_i' solange, bis die zu (2.3.19) entgegengesetzte Ungleichung erfüllt ist.

Gesucht werden also Vektoren, die vom Koordinatenursprung zum Punkt mit den Koordinaten α_i', $i=1, \ldots, s$, weisen und für die (2.3.19) nicht erfüllt ist. Offensichtlich ist die Wahrscheinlichkeit, daß Punkte mit den Koordinaten α_i' in einen Kegel, dessen Spitze im Kugelmittelpunkt liegt, fallen, proportional zu dem diesem Kegel entsprechenden Raumwinkel. Die Projektionen des gefundenen Vektors sind gleich

$$\frac{\alpha_i'}{\left(\sum\limits_{i=1}^{s} (\alpha_i')^2\right)^{1/2}}, \quad i = 1, \ldots, s.$$

Einzelheiten und gewisse Modifikationen des Algorithmus kann man in den Arbeiten von Buslenko u. a. [1], S. 134, und von Müller [2] finden. Die Effektivität der beschriebenen Methode ist offenbar gleich dem Volumen der s-dimensionalen Kugel mit dem Radius 1/2. Bei großem s ist dieses Volumen klein. So ergeben für $s = 6$ nur etwa 2% aller Versuche ein gewünschtes Resultat. In diesen Fällen ist der Übergang zu Polarkoordinaten und die Modellierung der entsprechenden Winkelverteilungen zweckmäßig.

Man könnte die Aufzählung verschiedener Verfahren zur Modellierung konkreter Verteilungen noch fortsetzen. Wir haben uns aber auf charakteristische Verfahren zur Modellierung der in praktischen Anwendungen am weitesten verbreiteten Verteilungen beschränkt.

Zum Abschluß seien noch einige Bemerkungen zur Modellierung zufälliger Prozesse angefügt.

Es wurden bereits Fragen der Modellierung von Realisierungen eines Poissonschen Prozesses untersucht. Dieser gehört zu den diskreten Markoffschen Prozessen, deren Modellierung in einer Folge von Modellierungen von Zufallsgrößen besteht, die die Übergangswahrscheinlichkeit bestimmen. Bezüglich der Modellierung allgemeinerer zufälliger Prozesse kann man nur folgendes sagen. Für die Erzeugung der Realisierung eines zufälligen Prozesses müssen eigentlich alle möglichen (darunter auch die unendlichdimensionalen) gemeinsamen Verteilungen ihrer Werte modelliert werden. Praktisch können wir allerdings keine Realisierung eines unendlichdimensionalen Zufallsvektors erhalten und sind daher gezwungen, uns auf einen endlichdimensionalen Vektor mit genügend vielen Komponenten zu beschränken. Der dadurch hervorgerufene Fehler hängt vom Charakter der konkreten Aufgabe ab und erfordert in jedem Fall eine besondere Untersuchung.

Sollen durch die Modellierung gewisse Mittelwerte über die Trajektorien des Prozesses berechnet werden, so ist der Übergang zu endlichdimensionalen Verteilungen äquivalent der näherungsweisen Ersetzung eines unendlichfachen Integrals durch ein gewöhnliches mehrfaches Integral. Der Fehler hängt dabei von der Art der zu integrierenden Funktion ab und nimmt ab, je weniger die Funktion von den Variablen mit hohem Index beeinflußt wird. Nach Durchführung der entsprechenden Untersuchung und Bestimmung der Dimension des Vektors, der der Modellierung zugrunde gelegt wird, kann die Aufgabe prinzipiell mit einer der oben beschriebenen Methoden gelöst werden. Dabei kann der Algorithmus der Modellierung natürlich sehr kompliziert sein.

Von bedeutendem Interesse ist die Ausarbeitung spezieller Verfahren zur Modellierung einzelner Klassen von zufälligen Prozessen. In dieser Richtung gibt es gegenwärtig nur vereinzelte Resultate, die hauptsächlich stationäre Gaußsche Prozesse betreffen.

Wir gehen kurz auf einige dieser Resultate ein. Weil die endlichdimensionalen Verteilungen eines Gaußschen Prozesses Normalverteilungen sind, kann man das schon beschriebene Verfahren zur Modellierung eines normalverteilten Vektors benutzen. Ein solcher Weg scheint unbequem nicht nur wegen der numerischen Schwierigkeiten bei der Modellierung eines Vektors hoher Dimension, sondern auch, weil es nicht immer möglich ist, die erforderliche Dimension ohne vorbereitende Rechnungen richtig zu bestimmen. Reicht die gewählte Dimension nicht aus, so ist eine zusätzliche Rechnung notwendig, die mit der Inversion der entsprechenden Matrix hoher Ordnung verknüpft ist. Manchmal ist es möglich, diese Schwierigkeiten zu umgehen. Beispielsweise kann man bei der Konstruktion der Trajektorien einer Brownschen Bewegung (Wiener [1], Gelfand, Frolow und Tschenzow [1]) die Dimension der zu modellierenden Zufallsvektoren schrittweise ohne zusätzliche Rechnungen erhöhen.

Es sei $\xi(t)$ die Trajektorie eines Brownschen Teilchens für $0 \leqq t \leqq 1$ mit $\xi(0)=0$. Die Trajektorie $\xi(t)$ kann durch einen Polygonzug approximiert werden, dessen Ecken mit den Punkten $t_i = i/2^n$, $i = 0, 1, \ldots, 2^n$, zusammenfallen. Zunächst ist zu bemerken, daß für $n = 1$ die Zufallsgröße $\xi(1)$ normalverteilt ist mit dem Erwartungswert Null (wegen $\xi(0) = 0$) und der Dispersion Eins. Ihre Realisierung

kann durch eine der früher untersuchten Methoden erhalten werden. Eine Vergrößerung der Zahl n zieht keine Neuberechnung der früher erhaltenen Realisierungen nach sich. Für $n = 2$ ist $\xi(1/2)$ normalverteilt mit dem Erwartungswert $\frac{\xi(0) + \xi(1)}{2}$ und der Dispersion 1/4. Wurden allgemeiner die Realisierungen $\xi(t')$ und $\xi(t'')$, $t'' > t'$, erhalten, so ist $\xi\left(\frac{t' + t''}{2}\right)$ normalverteilt mit dem Erwartungswert $\frac{\xi(t') + \xi(t'')}{2}$ und der Dispersion $(t'' - t')/2$. Dadurch wird es möglich, das Intervall [0, 1] sukzessive weiter zu halbieren und einen Polygonzug mit entsprechender Eckenzahl zu konstruieren. Dabei kann die Konvergenz des uns interessierenden Funktionals auf den Trajektorien gegen seinen Grenzwert festgestellt werden. In der Arbeit von Gelfand u. a. [1] wird ein Beispiel für die Fehlerabschätzung bei einem konkreten Funktional durchgerechnet.

Rekursive Prozeduren zur Modellierung einiger Klassen von Prozessen werden in der Arbeit von Sragowitsch [1] angegeben. Besitzt beispielsweise ein stationärer Gaußscher Prozeß eine echt gebrochene rationale Spektraldichte[1])

$$g(\lambda) = \left|\frac{A(\lambda)}{B(\lambda)}\right|$$

mit

$$\left.\begin{aligned} A(\lambda) &= A_r\lambda^r + \cdots + A_1\lambda + A_0 = A_r \prod_{k=1}^{r} (\lambda - \tilde{\lambda}_k)\,, \\ B(\lambda) &= \lambda^s + B_{s-1}\lambda^{s-1} + \cdots + B_1\lambda + B_0 = \prod_{k=1}^{s} (\lambda - \lambda_k)\,, \\ &\quad r < s\,, \qquad \operatorname{Im} \lambda_k > 0\,, \end{aligned}\right\} \tag{2.3.20}$$

so wählt man einige Δt aus und betrachtet die Werte $\xi_k = \xi(t_k)$ des Prozesses in den Punkten $t_k = k\Delta t$. Bei genügend kleinem Δt nähert die Realisierung einer stationären Folge der ξ_k eine Realisierung des zufälligen Prozesses beliebig gut an. Für diese Folge soll die Spektraldichte die Form

$$f(\lambda) = |\tilde{A}(e^{i\lambda})/\tilde{B}(e^{i\lambda})|$$

haben. Unter der Annahme reeller Prozesse, wobei $\tilde{A}(\lambda)$ und $\tilde{B}(\lambda)$ Polynome der Gestalt (2.3.20) mit reellen Koeffizienten sind, kann man zeigen: Sind die Nullstellen von $\tilde{B}(\lambda)$ betragsmäßig kleiner als 1, so genügen die ξ_k der Differenzengleichung

$$\xi_{s+l} + \tilde{B}_{s-1}\xi_{s+l-1} + \cdots + \tilde{B}_0\xi_l = \tilde{A}_r\eta_{l+r} + \cdots + \tilde{A}_1\eta_{l+1} + \tilde{A}_0\eta_l\,, \tag{2.3.21}$$

wobei die $\tilde{A}_i$, $i = 0, \ldots, r$, und $\tilde{B}_j$, $j = 1, \ldots, s-1$, die Koeffizienten der Polynome $\tilde{A}$ bzw. $\tilde{B}$ und die η unabhängige mit dem Erwartungswert Null und der

---

[1]) Die Spektraldichte $g(\lambda)$ eines stationären zufälligen Prozesses ist mit seiner Korrelationsfunktion $R(t) = M(\xi(t)\,\xi(t+\tau))$ durch folgende Beziehung

$$R(t) = \int_{-\infty}^{+\infty} e^{i\lambda t}\, g(\lambda)\, d\lambda$$

verknüpft.

Dispersion 2π normalverteilte Größen sind. Es gilt auch die umgekehrte Behauptung: Genügt die stationäre Folge der ξ_k der Differenzengleichung (2.3.21) und sind die Nullstellen des Polynoms $\lambda^s + \sum_{k=0}^{s-1} \tilde{B}_k \lambda^k$ betragsmäßig kleiner als 1, so existiert eine Spektraldichte dieser Folge von der Gestalt $f(\lambda)$.

Wenn also $K(\tau) = ce^{-\beta|\tau|}$, $\beta > 0$, die Korrelationsfunktion des Prozesses ist, so ergibt sich für die entsprechende Spektraldichte

$$g(\lambda) = \frac{A}{\lambda^2 + \beta^2} \quad \text{mit} \quad A = \frac{\beta}{\pi} c .$$

Die Korrelationsfunktion der Folge der ξ_k ist

$$R(k) = ce^{-\beta|k|\Delta t} = c\varrho^{|k|}$$

mit

$$\varrho = e^{-\beta\Delta t}, \quad f(\lambda) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} R(k) e^{-i\lambda k} .$$

Es ist leicht zu sehen, daß

$$f(\lambda) = \frac{B^2}{|e^{i\lambda} - \varrho|^2}, \quad B^2 = \frac{1 - \varrho^2}{2\pi} c$$

ist und die Gleichung (2.3.21) die Gestalt

$$\xi_{k+1} - \varrho\xi_k = B\eta_k \tag{2.3.22}$$

annimmt. Die Modellierung besteht in der Erzeugung unabhängiger Realisierungen von normalverteilten Zufallsgrößen η_k mit den Parametern $(0, 2\pi)$ und Berechnungen nach der Rekursionsformel (2.3.22). Ist die Spektraldichte in einfache Brüche zerlegbar, so kann die Beziehung (2.3.21) durch eine Reihe von Rekursionsformeln der Art (2.3.22) ersetzt werden. In der Arbeit von Sragowitsch werden auch Rekursionsbeziehungen für die Modellierung von Diffusionsprozessen angegeben.

Ein anderer Weg zur Modellierung Gaußscher Prozesse wird in der Arbeit von Piranaschwili [1] untersucht. Dort wird die Aufgabe betrachtet, einen Prozeß $\xi(t, \omega)$, $a \leqq t \leqq b$, mit gegebener eindimensionaler Verteilungsfunktion $f(x, t)$ und Korrelationsfunktion $R(t, s)$ zu modellieren. Dabei ist $y = f(x, t)$ absolut stetig und $x = f_t^{-1}(y)$ die zu f inverse Funktion für festes t.

Für eine umfangreiche Klasse Gaußscher Prozesse gilt die Darstellung

$$\xi(t, \omega) = f_t^{-1}(\Phi(\eta(t, \omega))), \qquad a \leqq t \leqq b, \tag{2.3.23}$$

mit

$$\eta(t, \omega) = \sum_{k=1}^{\infty} z_k(\omega) \frac{\varphi_k(t)}{\sqrt{\lambda_k}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du .$$

Die $z_k(\omega)$ sind unabhängige, mit den Parametern (0, 1) normalverteilte Zufallsgrößen, λ_k und φ_k sind die Eigenwerte und Eigenfunktionen der Integralgleichung

$$\varphi(t) = \lambda \int_a^b r(t, s)\, \varphi(s)\, ds\,.$$

Der Kern dieser Gleichung ist mit der Korrelationsfunktion $R(t, s)$ des Prozesses $\xi(t, \omega)$ durch eine ziemlich komplizierte Beziehung verknüpft, und zwar ist $r(t, s)$ Wurzel der Gleichung

$$\psi\big(r(t, s)\big) = 0 \tag{2.3.24}$$

mit

$$\psi(r) = \frac{2}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\infty} f_t^{-1}\big(\Phi(\tau_1)\big)\, f_s^{-1}\big(\Phi(\tau_2)\big) \cdot \exp\left\{-\frac{\tau_1^2 - 2r\tau_1\tau_2 + \tau_2^2}{2(1-r^2)}\right\} d\tau_1 d\tau_2$$
$$- m(t)\, m(s) - R(t, s)\, \sigma(t)\, \sigma(s)\,. \tag{2.3.25}$$

Dabei sind $m(t)$ und $\sigma^2(t)$ der Erwartungswert und die Dispersion des Prozesses $\xi(t, \omega)$.

Die Darstellung (2.3.23) ist möglich, wenn die Gleichung (2.3.24) eine solche Lösung $r(t, s)$ besitzt, daß alle Eigenwerte der oben genannten Integralgleichung nichtnegativ sind. Diese Darstellung gestattet es dann, den Gaußschen Prozeß mit eindimensionaler Dichte und Korrelationsfunktion zu modellieren. Aber die vorausgehende Arbeit, die notwendig ist, um diese Darstellung zu erhalten, kann recht kompliziert sein.

Diese Resultate, die wir ohne ausführliche Erläuterungen anführten, charakterisieren in genügendem Maß die Spezifik der Aufgaben, die bei der Modellierung zufälliger Prozesse entstehen. Der Leser, der sich speziell für diese Fragen interessiert, findet eine Reihe von ergänzenden Ausführungen und Algorithmen auch in der Arbeit von Franklin [2].

III. SIMULATION

§ 1. Durchgang von Strahlung durch Materie

Wenn man die in den beiden vorhergehenden Kapiteln erworbenen Kenntnisse benutzt, kann man eine Reihe praktischer Aufgaben durch Simulation des stochastischen Modells der betreffenden realen Erscheinung lösen. In diesem Kapitel werden wir dabei absichtlich formale Schlußweisen oder die Untersuchung von Gleichungen der einen oder anderen Erscheinung vermeiden. Als wesentlich soll eine Besonderheit der Monte-Carlo-Methode hervorgehoben werden, die in folgendem besteht. Ihre Anwendung auf Probleme des Strahlendurchgangs oder der Massenbedienung erfordert eigentlich nicht die Formulierung von Integrodifferentialgleichungen oder Differentialgleichungssystemen. In der Praxis kann die Situation entstehen, daß die Formulierung solcher Gleichungen erhebliche Schwierigkeiten bereitet und es wünschenswert ist, die Aufgabe ohne analytische Untersuchung des stochastischen Modells zu lösen. Allerdings gestattet die Benutzung des analytischen Apparates, effektivere Algorithmen zur Lösung der Aufgaben aufzustellen, wie wir im weiteren sehen werden.

Wie bereits bemerkt, ist die Entstehung der Monte-Carlo-Methode eng mit Aufgaben zur Berechnung des Strahlendurchgangs durch Materie verknüpft. Viele theoretische Fragen der Methode entstanden und entwickelten sich nämlich in Verbindung mit diesen Aufgaben. Daher scheint es nützlich, zuerst stochastische Modelle des Strahlendurchgangs durch Materie darzustellen.

Notwendigerweise wird diese Darlegung aber unvollständig sein, weil der Autor sich außerstande sieht, in größerem Umfang Fragen der Neutronen- oder γ-Strahlenphysik zu behandeln. Im wesentlichen wird das Schema der Modellierung dargestellt, das allerdings den physikalischen Charakter der Erscheinung widerspiegelt. Wir werden den Begriff „Teilchen“ als Sammelbezeichnung gebrauchen, wenn wir Neutronen, γ-Quanten oder Protonen meinen. Zunächst stellen wir das Schema der Modellierung in allgemeiner und idealisierter Form dar. Es sei ein gewisses Volumen $\mathcal{D}$ gegeben, das von Atomen eines Stoffes erfüllt und von Vakuum umgeben ist. Das Teilchen wird durch seine Lage im Raum $\boldsymbol{r} = (x, y, z)$, die Zeitkoordinate t, die Bewegungsrichtung $\boldsymbol{\Omega} = (u, v, w)$ mit $\boldsymbol{\Omega}^2 = 1$ und die Geschwindigkeit v (oder Energie E) charakterisiert (im Fall der γ-Quanten kann nur $E = h\nu$ benutzt werden). Die das Teilchen charakterisierenden Größen werden also durch einen Punkt in einem siebendimensionalen Phasenraum dargestellt. Wir nehmen dabei an (Frolow und Tschenzow [3]), daß

a) die Konzentration der Teilchen im Körper so klein ist, daß man ihre gegenseitige Wechselwirkung vernachlässigen kann;

b) die Wechselwirkung der Teilchen mit der Materie derart ist, daß die Wahrscheinlichkeit einer bestimmten Wechselwirkung nicht von dem vorhergehenden Verhalten der Teilchen abhängt;

c) die Materie sich im stationären Zustand befindet, d. h., die Dichte des Stoffes und Zusammensetzung sich nicht mit der Zeit ändern.

Unter diesen Voraussetzungen setzt sich der Prozeß aus den unabhängigen „Geschichten" der einzelnen Teilchen zusammen. Durch Verfolgung dieser Geschichten kann man verschiedene Charakteristiken des Prozesses des Strahlendurchgangs erhalten. Die Geschichte eines Teilchens beginnt mit seiner Entstehung. Im Gebiet T des Phasenraumes der Teilchenkoordinaten ist die Verteilungsfunktion $f(\boldsymbol{r}, \boldsymbol{\Omega}, E, t)$ der Quellen gegeben. Durch Modellierung dieser Verteilung im Zeitmoment t_0 erhalten wir die Anfangskoordinaten $\boldsymbol{r}_0$, $\boldsymbol{\Omega}_0$, E_0, t_0 des Teilchens. Dann bewegt sich das Teilchen geradlinig und gleichförmig weiter bis zum Zusammenstoß mit einem Atomkern der Substanz (bzw. auch mit einem Elektron bei γ-Quanten). Die Wahrscheinlichkeit eines Stoßes auf dem Weg dl beträgt $\Sigma(\boldsymbol{r}, E)\,dl$. Die Größe $\Sigma(\boldsymbol{r}, E)$ wird für den gegebenen Stoff und den gegebenen Teilchentyp experimentell bestimmt (manchmal ist auch eine theoretische Berechnung möglich). Dabei ist $\Sigma(\boldsymbol{r}, E) = \sum_{i=1}^{m} \varrho_i(\boldsymbol{r})\,\sigma_i(E)$. Hierbei ist $\varrho_i(\boldsymbol{r})$ die Dichte des Elements mit der Nummer i (Anzahl der Kerne bzw. Elektronen in 1 cm^3), $\sigma_i(E)$ der mikroskopische Wirkungsquerschnitt für ein Element mit der Nummer i und m die Anzahl von Isotopen der chemischen Elemente, die bei der Berechnung zu berücksichtigen sind. Für Neutronen ist $\sigma_i = d_i/n$, wobei n die Anzahl von Atomen in einer einatomigen Schicht mit der Fläche 1 cm^2 und d_i die Anzahl von Neutronen ist, die mit dem gegebenen Element unter der Bedingung in Wechselwirkung treten, daß n Neutronen senkrecht zur Schicht einfallen. $\Sigma(\boldsymbol{r}, E)$ heißt auch totaler makroskopischer Wirkungsquerschnitt für den gegebenen Stoff. Die Entfernung l, die das Teilchen in der gegebenen Bewegungsrichtung $\boldsymbol{\Omega}$ ohne Wechselwirkung zurücklegt, ist also nach einem Exponentialgesetz mit dem Parameter $\Sigma(\boldsymbol{r}, E)$ verteilt:

$$H(l) = 1 - \exp\left\{-\int_0^l \Sigma(\boldsymbol{r}_0 + l\boldsymbol{\Omega}, E)\,dl\right\}, \tag{3.1.1}$$

wobei $\boldsymbol{r}_0$ die Anfangslage des Teilchens ist. Durch Modellierung dieser Exponentialverteilung finden wir die Weglänge des Teilchens. Es kann sich dabei herausstellen, daß das Gebiet $\mathcal{D}$ von dem Teilchen ohne Wechselwirkung verlassen wird. Dann muß seine Geschichte offensichtlich als beendet betrachtet werden. Gehört aber der Punkt mit den Koordinaten

$$x_1 = x_0 + u_0 l\,, \qquad y_1 = y_0 + v_0 l\,, \qquad z_1 = z_0 + w_0 l \tag{3.1.2}$$

zum Gebiet $\mathcal{D}$, so sind nun mit verschiedenen Wahrscheinlichkeiten verschiedene Arten der Wechselwirkung möglich (ein Neutron kann mit Kernen verschiedener Elemente zusammenstoßen usw.). Wenn wir diese Wahrscheinlichkeiten kennen, dann können wir die entsprechende diskrete Verteilung modellieren und die konkrete Form der stattgefundenen Wechselwirkung bestimmen. Im einfachsten Fall gibt es drei Typen von Wechselwirkungen: Streuung, Absorption und Zerfall

(Bildung neuer Teilchen). Bei der Streuung ändern sich, grob gesagt, die Bewegungsrichtung des Teilchens und seine Energie. Bei der Absorption wird das Teilchen durch den Stoff aufgenommen, und seine Geschichte ist beendet. Beim Zerfall entstehen ein oder mehrere neue Teilchen, d. h., es bildet sich eine neue (sekundäre) Quelle.

1. Streuung. Die Streuungswahrscheinlichkeit ist gleich $\Sigma_{\text{Str}}/\Sigma$. Dabei wird der makroskopische Streuquerschnitt Σ_{Str} analog zum totalen Wirkungsquerschnitt Σ bestimmt. Σ_{Str} stellt die Summe aller möglichen Arten von Streuung dar. Für jeden konkreten Streuungstyp wird die Wahrscheinlichkeitsdichte dafür angegeben, daß sich die Bewegungsrichtung von $\boldsymbol{\Omega}$ in $\boldsymbol{\Omega}'$ und die Energie von E in E' ändert. Diese Dichte $g(\boldsymbol{\Omega}, E \to \boldsymbol{\Omega}', E')$ heißt Indikatrix der Streuung. Ihre Modellierung gibt die Möglichkeit, die neue Bewegungsrichtung und die neue Energie des Teilchens zu bestimmen. Von neuem wird dann wieder der Weg bestimmt, den das Teilchen gemäß (3.1.1) zurücklegt, d. h., die Berechnung der Geschichte des Teilchens besteht aus Operationen vom gleichen Typ. Sie endet, wenn das Teilchen das Gebiet $\mathcal{D}$ verläßt, absorbiert wird oder zerfällt. Manchmal endet die Rechnung auch, wenn die Energie des Teilchens einen gewissen unteren zulässigen Wert $E_{\min}$ unterschreitet.

2. Absorption. Ihre Wahrscheinlichkeit beträgt $\Sigma_{\text{Abs}}/\Sigma$. Die physikalische Natur der Absorption kann verschieden sein, aber für uns ist nur wesentlich, daß mit der Absorption die Geschichte des Teilchens beendet ist.

3. Zerfall. Die Wahrscheinlichkeit des Zerfalls wird durch die Größe $\Sigma_{\text{Zerf}}/\Sigma$ bestimmt. Als Ergebnis des Zerfalls entstehen ein oder mehrere neue Teilchen. Die Verteilungsfunktion der Phasenkoordinaten der entstehenden Teilchen ist anzugeben; durch die Modellierung werden diese Koordinaten dann berechnet. Bei der Entstehung mehrerer Teilchen verzweigt sich der Prozeß, d. h., es müssen nun die Geschichten jedes entstandenen Teilchens modelliert werden. Wenn die sekundären Teilchen von neuem einen Zerfallsprozeß hervorrufen können, entsteht ein „Baum“, von dem alle „Zweige“ betrachtet werden müssen. Spezielle Methoden zur Untersuchung solcher „Bäume“ sind in der Literatur beschrieben (z. B. Golenko [1], Frolow [1]). Es ist noch zu bemerken, daß die Bestimmung der Zeitkoordinate t bei bekannter Bewegungsgeschwindigkeit und zurückgelegtem Weg keine Mühe macht.

So stellt sich in groben Zügen das Bild dar, das bei der unmittelbaren Modellierung des Strahlendurchgangs durch Materie entsteht. Die Resultate der Rechnung sind gewöhnlich einige Mittelwerte, beispielsweise die mittlere Anzahl der Teilchen, die das Gebiet $\mathcal{D}$ verlassen, die mittlere Anzahl der absorbierten Teilchen usw. Komplizierter in ihrer Realisierung sind Aufgaben, in denen die Verteilung der Teilchen in Abhängigkeit von irgendwelchen Charakteristiken (Energie, Richtung usw.) zu bestimmen ist.

Wir gehen ausführlicher auf konkrete Beispiele ein.

Beispiel 1. **Durchgang von γ-Quanten durch Materie.** Das Gebiet $\mathcal{D}$ sei eine Schicht eines homogenen Stoffes, die in zwei Dimensionen als beliebig (unendlich) groß angenommen wird und die eine endliche Dicke H in Richtung der z-Achse

hat. Die Strahlenquelle soll punktförmig und im Innern der Schicht gelegen sein, ihre Strahlung sei isotrop und monochromatisch. Mit anderen Worten, die γ-Quanten entstehen alle in demselben Punkt $\boldsymbol{r}_0$ mit der festen Energie E_0 und besitzen die gleiche Wahrscheinlichkeit, sich in beliebiger Richtung zu bewegen (Abb. 3.1). Die Wirkungsquerschnitte können in Form von Tabellen gegeben sein (der totale Querschnitt $\Sigma(E)$ und der Streuquerschnitt $\Sigma_{\text{Str}}(E)$). Zwischenwerte findet man durch Interpolation. (Die verschiedenen Prozesse der Wechselwirkung von γ-Quanten mit der Materie zerlegen wir in zwei Gruppen — Streuprozesse und Absorptionsprozesse. Wenig wahrscheinliche Prozesse lassen wir unberücksichtigt, wobei wir insbesondere nur die Streuung der γ-Quanten an den Elektronen des Stoffes mitrechnen.) Der Algorithmus ist vollständig bestimmt durch die Angabe

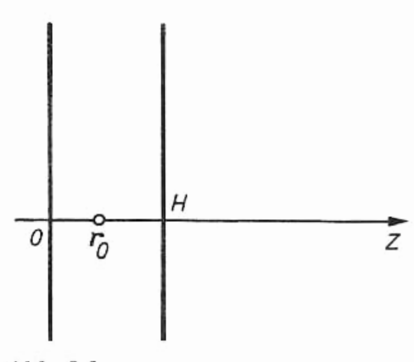


Abb. 3.1

der Regeln des Übergangs vom Punkt $\boldsymbol{x}_n = (\boldsymbol{r}_n, \boldsymbol{\Omega}_n, E_n)$ zum Punkt $\boldsymbol{x}_{n+1} = (\boldsymbol{r}_{n+1}, \boldsymbol{\Omega}_{n+1}, E_{n+1})$ des Phasenraumes und besteht in folgendem:

1. Aus E_n ergibt sich $\Sigma(E_n)$ und $\Sigma_{\text{Str}}(E_n)$.

2. Wir ermitteln $l_{n+1} = -\log\alpha/\Sigma(E_n)$ als Länge, die das γ-Quant bis zur Wechselwirkung mit den Elektronen des Stoffes zurücklegt.

3. $x_{n+1} = x_n + l_{n+1}u_n$, $y_{n+1} = y_n + l_{n+1}v_n$, $z_{n+1} = z_n + l_{n+1}w_n$. Fällt $z_{n+1} > H$ oder $z_{n+1} < 0$ aus, so wird der Austritt des γ-Quants aus der Platte registriert.

4. Ist $0 \leqq z_{n+1} \leqq H$, so ist die Art der Wechselwirkung zu bestimmen. Mit der Wahrscheinlichkeit $\Sigma_{\text{Str}}/\Sigma$ findet eine Streuung statt, mit der Wahrscheinlichkeit $1 - \Sigma_{\text{Str}}/\Sigma$ Absorption, und damit endet die Geschichte dieses γ-Quants.

5. Die Dichte der Wahrscheinlichkeitsverteilung von E_{n+1}, gemessen in Einheiten mc^2 (m Elektronenruhemasse, c Lichtgeschwindigkeit) wird durch die aus relativistisch invarianten Beziehungen ableitbare Formel von KLEIN-NISHINA-TAMM bestimmt (LEIPUNSKI u. a. [1]):

$$\varphi(E_{n+1}, E_n) = \begin{cases} C\left(\dfrac{E_{n+1}}{E_n}\right)^2 \left[\dfrac{E_{n+1}}{E_n} + \dfrac{E_n}{E_{n+1}} + 2\left(\dfrac{1}{E_n} - \dfrac{1}{E_{n+1}}\right) + \left(\dfrac{1}{E_n} - \dfrac{1}{E_{n+1}}\right)^2\right] \\ \qquad\qquad \text{für } E_n \geqq E_{n+1} \geqq E_n/(2E_{n+1})\,, \\ 0 \qquad\qquad \text{sonst.} \end{cases} \tag{3.1.3}$$

Dabei ist C eine von E_n abhängige Normierungskonstante. In der Arbeit von ISHII u. a. [1] wird eine bedeutend einfachere Formel angegeben, durch die die Werte von E_{n+1} mit einem Fehler von etwa 5% in einem großen Bereich der Energiewerte E_n berechnet werden:

$$E_{n+1} = \frac{E_n}{1 + S\alpha + 2(E_n - S)\,\alpha^3} \quad \text{mit} \quad S = \frac{E_n}{1 + 0{,}5625\,E_n}.$$

6. Der Kosinus des Streuwinkels wird nach der Formel

$$\cos\vartheta_{n+1} = 1 + \frac{1}{E_n} - \frac{1}{E_{n+1}}$$

bestimmt. Die Abweichung von der ursprünglichen Richtung ist eindeutig durch die Energie bestimmt. Der Azimutwinkel ist gleichverteilt.

7. Durch den Winkel ϑ_{n+1} und den isotrop verteilten Azimutwinkel wird die neue Richtung $\boldsymbol{\Omega}_{n+1}$ nach der Streuung bestimmt (Abb. 3.2). Die abschließende Formel kann man beispielsweise in dem Handbuch von BUSLENKO u. a. [1] finden:

$$\left.\begin{aligned} u_{n+1} &= \frac{bcw_n u_n - bdv_n}{(1 - w_n^2)^{1/2}} + au_n \\ v_{n+1} &= \frac{bcw_n v_n - bdu_n}{(1 - w_n^2)^{1/2}} + av_n \\ w_{n+1} &= -\,bc(1 - w_n^2)^{1/2} + aw_n \end{aligned}\right\} \tag{3.1.4}$$

mit $a = \cos\vartheta_{n+1}$, $b = \sin\vartheta_{n+1}$, $c = \cos(2\pi\alpha)$, $d = \sin(2\pi\alpha)$. Damit ist der Algorithmus vollständig beschrieben.

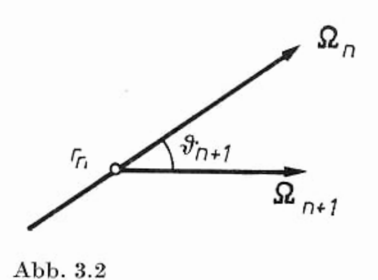


Abb. 3.2

Tabellen der Wirkungsquerschnitte für γ-Quanten haben verhältnismäßig geringen Umfang. Sie werden z. B. in dem Buch von LEIPUNSKI u. a. [1] angegeben. Die Winkel- und Energiecharakteristiken können durch verhältnismäßig einfache theoretische Formeln erhalten werden. Anders ist das im Fall der Neutronen. Die Modellierung des Durchgangs von Neutronen durch Materie erfordert die Speicherung einer großen Anzahl von Konstanten. Dabei ist die Winkelverteilung, die experimentell bestimmt werden muß, nicht immer genau genug bekannt. Man kann annehmen, daß die Mittelwerte, die man gewöhnlich als Resultat der Rechnung erhält, verhältnismäßig wenig von Fehlern in der Winkelverteilung oder

anderen weniger wahrscheinlichen Effekten abhängen. Allerdings erfordern diese Überlegungen in konkreten Fällen zusätzliche Begründungen, zusätzliche Rechnungen oder Experimente. Bei der beschriebenen Modellierungsmethode gibt es keine prinzipiellen Hindernisse, beliebige Mittelwerte des Prozesses zu bestimmen. Jedoch läßt sich beispielsweise die mittlere Anzahl von Teilchen, die in ein sehr kleines Volumen des Phasenraums fallen, praktisch nicht bestimmen. Denn die Wahrscheinlichkeit, dieses Volumen zu treffen, ist klein, und daher sind sehr viele Versuche notwendig, um wenigstens ein in dieses Volumen fallendes Teilchen zu erhalten. Für die Berechnung kleiner Wahrscheinlichkeiten empfiehlt es sich, „fiktive" Modelle aufzustellen. Als einfaches Beispiel eines fiktiven Modells kann das folgende dienen.

γ-Quanten dringen in eine Schicht ein, in der sie stark absorbiert werden. Uns interessiert nun die Wahrscheinlichkeit des Durchgangs. Wir nehmen an, daß sich auf derselben Trajektorie nicht ein γ-Quant, sondern eine Gruppe von N Teilchen bewegt. In einem Punkt der Wechselwirkung werden im Mittel $N\Sigma_{\text{Str}}/\Sigma$ Teilchen gestreut und die restlichen absorbiert. Wir führen nun für γ-Quanten ein sogenanntes „statistisches Gewicht" P_n ein, das bei der Entstehung des Teilchens den Wert $P_0 = 1$ hat und im Punkt der n-ten Wechselwirkung den Wert $P_n = P_{n-1}\Sigma_{\text{Str}}(E_n)/\Sigma(E_n)$ erhält. Beim Austritt aus der Schicht zählt man nicht die Menge der austretenden γ-Quanten, sondern die Summe der Gewichte der austretenden Teilchen (die austretenden Teilchen können verschiedenes Gewicht haben). Alle γ-Quanten liefern so einen Beitrag in der Summe der austretenden Teilchen, und von vornherein ist zu erwarten, daß diese Summe recht genau berechnet werden kann. Unsere Erwägungen bedürften jedoch noch eines strengen Beweises. Die Benutzung ähnlicher Verfahren ohne strenge Begründung kann auch zu groben Fehlern führen.

Beispiel 2. *Modellierung des Kernkaskadenprozesses.* Ein verhältnismäßig einfaches Beispiel für ein Modell, in dem das primäre Teilchen nach Wechselwirkung mit einem Atomkern die Bildung sekundärer Teilchen hervorruft, ist das Modell des Kaskadenprozesses. Es gibt verschiedene theoretische Modelle des Atomkerns, die in der Regel einige unbestimmte Parameter enthalten. Es ist anzunehmen, daß das Modell die Kernreaktionen befriedigend beschreibt, wenn es nicht allzuviele Parameter sind und diese so gewählt wurden, daß eine gute Übereinstimmung mit dem Experiment gewährleistet ist. Für die Auswahl der Parameter müssen umfangreiche Rechnungen durchgeführt werden. Ein Modell, dessen Parameter bestimmt sind, wird dann bei der Lösung verschiedener Aufgaben benutzt, die mit Kernreaktionen in Verbindung stehen (z. B. bei Berechnungen, die den Strahlenschutz betreffen).

Es muß gesagt werden, daß die Monte-Carlo-Methode gegenwärtig die einzige Methode ist, mit der man Charakteristiken von Kernreaktionen unter genügend allgemeinen physikalischen Voraussetzungen erhalten kann. Unter Vernachlässigung des physikalischen Aspekts (vgl. dazu Bernardini, Booth und Lindenbaum [1]) werden wir die Modellierung von Prozessen mit der Monte-Carlo-Methode beschreiben, die bei der Bombardierung von Atomkernen mit sehr energiereichen Teilchen (Protonen, Neutronen u. a. mit mehr als 10^8 EV) vor sich gehen. Dabei gehen wir von einem der einfachsten Kernmodelle aus.

Wir nehmen an, daß der Kern eine Kugel vom Radius R darstellt (R ist ein Parameter des Modells). Die Nukleonen, aus denen der Kern besteht, bilden ein entartetes Fermi-Gas, und die Dichte der Kernmaterie ist im allgemeinen ungleichmäßig. Ein Teilchen, das in den Kern eindringt, stößt mit den Nukleonen des Kerns zusammen, wird an ihnen gestreut und kann durch Abgabe eines Teils seiner Energie ihren Austritt aus den Grenzen des Kerns hervorrufen. Dann ist es nötig, die Trajektorien des eingedrungenen Teilchens (das gestreut wird) und der sekundären Nukleonen (die gestoßen wurden) weiter zu verfolgen. Jedes Nukleon wird durch seinen Impuls $\boldsymbol{p}$ und seinen Ortsvektor $\boldsymbol{r} = (x, y, z)$ charakterisiert.

Die Modellierung besteht aus folgenden aufeinanderfolgenden Schritten:

1. Es werden die Eintrittskoordinaten des mit gegebener Energie einfallenden Teilchens bestimmt. Beispielsweise setzt man x_0 als mit der Dichte

$$\varphi(x_0) = \begin{cases} \dfrac{2x_0}{R^2} & \text{für} \quad x_0 \in [0, R]\,, \\ 0 & \text{sonst} \end{cases}$$

verteilte Zufallsgröße an und $z_0 = \sqrt{R^2 - x_0^2}$, $y_0 = 0$, $w_0 = -1$, $u_0 = v_0 = 0$ (der Koordinatenursprung fällt mit dem Mittelpunkt des Kerns zusammen).

2. Die (freie) Weglänge l eines beliebigen (gestreuten oder gestoßenen) Nukleons ohne Zusammenstoß mit anderen Nukleonen wird als exponentialverteilt angenommen. Der Parameter der Exponentialverteilung ist konstant, wenn die Dichte der Kernmaterie konstant ist. Ist das nicht der Fall, so kann das Verteilungsgesetz recht kompliziert sein. Zum Beispiel ist

$$\mathsf{P}(l \leqq x) = 1 - \exp\left\{-\int_0^x f(t)\,dt\right\}$$

mit

$$f(t) = c\,\{1 + \exp\left(a^{-1}\left[\left(\left(\sqrt{R^2 - r^2} - t\right)^2 + r^2\right)^{1/2} - R\right]\right)\}^{-1}\,. \qquad (3.1.5)$$

Dabei sind a und c Konstanten, c hängt von der Energie und Art der Nukleonen ab. Weiter ist $r = |\boldsymbol{r}|$, wobei $\boldsymbol{r}$ der Ortsvektor des Punktes ist, von dem das Nukleon herkommt. Die Modellierung einer Zufallsgröße mit einer derartigen Dichte kann durch eine in Kap. II, § 3, beschriebene Methode verwirklicht werden.

3. Wenn der $(n-1)$-te Zusammenstoß im Punkt $\boldsymbol{r}_{n-1}$ stattfand ($\boldsymbol{r}_0$ ist im ersten Schritt definiert) und der Impuls $\boldsymbol{p}_{n-1}$ bekannt ist, werden die Koordinaten des n-ten Zusammenstoßes in üblicher Weise durch

$$\boldsymbol{r}_n = \boldsymbol{r}_{n-1} + \boldsymbol{\Omega}_{n-1} l_n$$

bestimmt. Dabei muß die Bedingung $|\boldsymbol{r}_n| \leqq R$ geprüft werden. Wenn diese nicht erfüllt ist, bedeutet das, daß das Nukleon den Kern verlassen hat.

4. Weiter wird dann bestimmt, zwischen welchen Teilchen ein Zusammenstoß stattfindet. Ist das ankommende Teilchen ein Proton, so muß man die Wahrscheinlichkeit für den Zusammenstoß mit einem Neutron oder Proton kennen, um bestimmen zu können, mit welchem Teilchen der Zusammenstoß erfolgt (man vergleiche die Auswahl des Wechselwirkungstyps für γ-Quanten). Ist das an-

kommende Teilchen ein Neutron, so verfährt man analog, aber die entsprechenden Wahrscheinlichkeiten haben im allgemeinen andere Werte. Alle diese Wahrscheinlichkeiten müssen tabelliert oder durch eine empirische Formel gegeben sein. Sie hängen von den entsprechenden Geschwindigkeiten v der stoßenden Nukleonen ab. v ist dabei durch die Formel

$$v = \frac{|\boldsymbol{p}_1 - \boldsymbol{p}_2|}{(|\boldsymbol{p}_1 - \boldsymbol{p}_2|^2 + m^2)^{1/2}}$$

gegeben, wobei m die Ruhemasse von Proton und Neutron, $\boldsymbol{p}_1$ der Impuls des einfallenden Teilchens und $\boldsymbol{p}_2$ der Impuls des gestoßenen Teilchens ist. Man nimmt an, daß die Bewegungsrichtung des getroffenen Nukleons vor dem Zusammenstoß isotrop ist (gleichwahrscheinlich in allen Richtungen, vgl. Kap. II, § 3) und $p_2 = |\boldsymbol{p}_2|$ die Verteilungsfunktion

$$F(p_2) = \begin{cases} c_1 \int\limits_0^{p_2} x^2 \exp\left(\frac{x^2}{p_0}\right) dx & \text{für} \quad p_2 \in [0, E_\Phi], \\ 0 & \text{für} \quad p_2 < 0, \\ 1 & \text{für} \quad p_2 > E_\Phi \end{cases}$$

besitzt. Dabei sind p_0 und E_Φ gegebene Konstanten (E_Φ ist die Fermienergie).

5. Der Winkel zwischen den Bewegungsrichtungen der Teilchen nach dem Stoß wird durch die gegebene Streuungsindikatrix $f_1(\mu)$ bestimmt, wobei μ den Kosinus des Polwinkels bezüglich des Massenmittelpunktes bedeutet. Der Azimutwinkel wird als gleichmäßig verteilt angenommen. Die Gestalt von f_1 hängt von dem benutzten Wechselwirkungsmodell ab. Beispielsweise setzt man

$$f_1(\mu) = c_2\left(a_1\mu^4 + a_2(1 - \mu^2)^{3/2} + 1\right),$$

wobei c_2 eine Normierungskonstante ist. Die Größen a_1 und a_2 hängen von den Impulsen des einfallenden und des getroffenen Nukleons ab, genauer, von der Größe

$$\tau = \left[\frac{m}{2} \frac{|\boldsymbol{p}_1 - \boldsymbol{p}_2|}{m + (m^2 + |\boldsymbol{p}_1 - \boldsymbol{p}_2|^2)^{1/2}} + m^2\right]^{1/2} - m,$$

und sind tabelliert oder durch eine empirische Formel gegeben.

6. Weiter werden auf Grund der Erhaltungssätze (vgl. dazu Baraschenkow u. a. [1] und Bernardini u. a. [1]) die Energie und die Bewegungsrichtung (also der Impuls) beider Teilchen nach dem Stoß in einem mitbewegten Koordinatensystem bestimmt. Daraus kann dann die Bewegungsrichtung in einem festen Koordinatensystem bestimmt werden (vgl. Formel (3.1.4)). Setzt man $s = (|\boldsymbol{p}_1 - \boldsymbol{p}_2|^2 + m^2)^{1/2}$, so ist

$$E_1 = s - \frac{m(s^2 - m^2)}{2m(m + s)}(1 - \mu), \quad E_2 = m\left[1 + \frac{s^2 - m^2}{2m(m + s)}(1 - \mu)\right],$$

wobei E_1 die Energie des einfallenden Nukleons und E_2 die Energie des gestoßenen Nukleons nach dem Stoß bedeuten. Sind ϑ_1 und ϑ_2 die entsprechenden Winkel

der Abweichung von der ursprünglichen Bewegungsrichtung, so ergibt sich

$$\cos \vartheta_1 = (qq_1)^{-1} [E_1(s - m) + sm - m^2] ,$$

$$\cos \vartheta_2 = (qq_2)^{-1} (s + m) (E_2 - m) ,$$

wobei $q = |q_1 - q_2|$, $q_1 = (E_1^2 - m^2)^{1/2}$ und $q_2 = (E_2^2 - m^2)^{1/2}$ die neuen Impulse der Nukleonen sind. Der Azimutwinkel wird in den einfachsten Modellen für beide Teilchen als gleich $\varphi_{1,2} = \varphi_1 + \varphi_2$ angenommen, wobei φ_1 und φ_2 die Azimutwinkel der Teilchen vor dem Zusammenstoß sind. Weil man in Wirklichkeit das Modell als zylindersymmetrisch bezüglich der ursprünglichen Bewegungsrichtung des Nukleons annehmen kann, kann eine einfachere Formel als (3.1.4) benutzt werden. Wir gehen darauf aber nicht näher ein.

Das Pauliprinzip, das einen Zusammenstoß verbietet, der zu besetzten Zuständen führt, wird durch Prüfung der Bedingungen $E_1 - m > E_\Phi$ und $E_2 - m > E_\Phi$ berücksichtigt. Bei Nichterfüllung dieser Bedingungen wird der Zusammenstoß als nicht stattgefunden gewertet.

Obwohl die angegebenen Formeln ohne eine gebührende physikalische Begründung dem Leser überflüssig scheinen könnten, ist es nützlich, sich den Kompliziertheitsgrad der Algorithmen klarzumachen. Ohne diese Erkenntnis ist es schwer, von den verschiedenen Methoden zur Effektivitätserhöhung der Rechnung zu sprechen. Dabei ist noch zu bemerken, daß wir wie im Fall der γ-Quanten nur die einfachste Aufgabe betrachtet haben.

Wir wenden uns nun wieder der Modellierung des Kaskadenprozesses zu.

In Abb. 3.3 wird durch die fettgedruckte Linie der Weg des ankommenden Nukleons im Innern des Kerns schematisch dargestellt. Die punktierten Linien stellen die Bewegungsrichtungen der gestoßenen Nukleonen dar. Diese Nukleonen können ihrerseits mit anderen Nukleonen des Kerns zusammenstoßen und sekundäre gestoßene Nukleonen bilden, obwohl die Wahrscheinlichkeit dafür gering ist. Die Kaskade klingt sehr schnell ab. Ebenso wie das eingedrungene Nukleon können auch die gestoßenen Nukleonen den Kern nicht verlassen, wenn ihre Energie zu klein geworden ist (sie werden dann durch den Kern absorbiert). Im Prozeß der Modellierung wird gewöhnlich zuerst die Trajektorie des ankommenden Nukleons berechnet, während die Charakteristiken aller angestoßenen Nukleonen

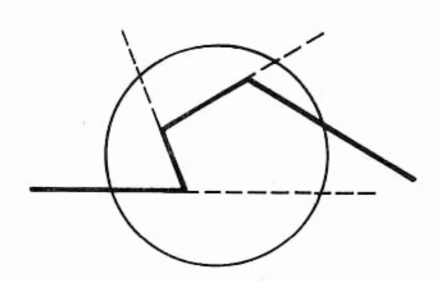


Abb. 3.3

in bestimmter Reihenfolge in der Maschine gespeichert werden. Dann berechnet man die Trajektorie des ersten gestoßenen Nukleons. Wenn es die Möglichkeit zur Bildung sekundärer gestoßener Nukleonen gibt, werden ihre Charakteristiken ebenfalls gespeichert. Nach Berechnung der Grundtrajektorie wendet man sich dann dem ersten sekundären Nukleon zu usw. So wird der ganze Baum der Trajektorien bis zum Ende jedes Zweiges berechnet. Im Speicher der EDVA wird dabei für die Speicherung der Charakteristiken jedes Zweiges ein entsprechender Platz bereitgestellt (ausführlich wird das bei GOLENKO [1] beschrieben).

Durch die Modellierung des Kernkaskadenprozesses kann man die Charakteristiken der den Kern verlassenden Nukleonen erhalten. Nachdem diese Nukleonen den Kern verlassen haben, bleibt er noch in einem angeregten Zustand. Es beginnt ein neues Stadium der Kernreaktion (das Verdampfungsstadium). Auf die Modellierung dieses Stadiums werden wir nicht eingehen (vgl. dazu z. B. BELJAJEW und ZARIZYNA [1]).

4. Einige Bemerkungen zur Realisierung von Algorithmen. Das Modell für den Strahlendurchgang durch Materie basiert auf einem stochastischen Modell dieses Vorgangs und verschiedenen, experimentell zu ermittelnden Konstanten, die in dieses Modell eingehen. Das Rechenprogramm besteht in der Regel aus unabhängigen Blöcken. Dieselben Blöcke können für die Lösung verschiedener Aufgaben benutzt werden. Es ist bequem, nicht irgendein vollständiges Programm zu haben, sondern ein System von Blöcken oder Prozeduren, aus denen man ein entsprechendes Programm zusammensetzen kann. Eine wichtige Rolle spielt die günstige Darstellung von Konstanten, die den Wechselwirkungsprozeß zwischen Strahlung und Materie charakterisieren. Daher ist es nützlich, in einem vorbereitenden Schritt empirische Formeln für die Wirkungsquerschnitte und die Streuungsindikatrix herzuleiten. Eine bequeme Darstellung solcher Abhängigkeiten kann die Rechenzeit erheblich verkürzen. Diese hängt wesentlich von der Qualität des Programms, von einer günstigen Speicherorganisation usw. ab. Haben z. B. nicht alle Angaben über die Wirkungsquerschnitte im Operativspeicher der EDVA Platz, so muß man die Aufgaben für ein gewisses Energieintervall in den Speicher bringen und die Trajektorien nicht nur eines Teilchens, sondern einer möglichst großen Anzahl von Teilchen in diesem Energieintervall verfolgen. Danach werden die Angaben über Wirkungsquerschnitte in den Speicher gebracht, die dem angrenzenden Energieintervall entsprechen usw.

Eine erhebliche Schwierigkeit kann die Aufstellung von Programmen für einen geometrischen Block bieten, in dem die freie Weglänge eines Teilchens zwischen zwei Punkten der Wechselwirkung mit der Materie bestimmt wird, wenn der Körper, in dem der Strahlendurchgang betrachtet wird, eine komplizierte geometrische Form hat oder aus Teilkörpern verschiedener Stoffe besteht. Für mit der Monte-Carlo-Methode zusammenhängende Aufgaben hat die genaue Formulierung und Untersuchung solcher Algorithmen eine wesentliche Bedeutung. Obwohl das vorliegende Buch hauptsächlich den statistischen Aspekten der Monte-Carlo-Methode gewidmet ist, gehen wir doch im Sinne eines Beispiels ausführlicher auf die Herstellung eines genügend allgemeinen geometrischen Blocks für die Aufgabenstellung des Strahlendurchgangs durch Materie ein (ERMAKOW und PROKOFJEWA [1], KUROPATENKO und OGIBIN [1]).

Die Voraussetzungen über die geometrische Gestalt des Körpers, durch den die Strahlung geht, bestehen in folgendem. Der Körper wird begrenzt durch eine geschlossene konvexe Fläche S. Im Innern von S befinden sich die geschlossenen Flächen S_i, die die aus verschiedenen Stoffen bestehenden Volumen V_i begrenzen. Für $i \neq j$ haben S_i und S_j keine gemeinsamen Punkte. Im übrigen ist ihre Lage innerhalb von S beliebig. Jeder der Stoffe wird durch den entsprechenden Index charakterisiert. Wir nennen eine Fläche, die sich innerhalb von S befindet, aber nicht innerhalb irgendeiner anderen Fläche eine Fläche ersten Ranges. Fläche vom Rang k heißt eine solche, die sich innerhalb irgendeiner Fläche vom Rang $k - 1$ befindet, aber nicht innerhalb einer Fläche von einem Rang größer als $k - 1$. Wir betrachten eine gewisse Fläche $S_i^{(m)}$ vom Rang m. Wir werden annehmen, daß der Index des Materials, aus dem diese Fläche besteht, auch dem Körper zukommt, den $S_i^{(m)}$ begrenzt und damit allen jenen Flächen vom Rang $m + 1$, die sich im Innern von $S_i^{(m)}$ befinden. Wenn man die Eigenschaften des Materials, aus dem S besteht, ähnlich den Eigenschaften des Vakuums vorgibt, dann kann man, wie leicht zu sehen ist, die Forderung der Konvexität der äußeren Fläche im wesentlichen weglassen. Berücksichtigt man weiter, daß verschiedene S_i sich nicht schneiden, aber doch beliebig dicht beieinander liegen können, so ist offensichtlich, daß die beschriebene geometrische Gestalt sehr allgemein ist.

Die Aufgabe des Blocks besteht nun in folgendem: Ein Teilchen bewegt sich vom Punkt mit den Koordinaten (x_n, y_n, z_n) in Richtung des Einheitsvektors (u_n, v_n, w_n). Es wird angenommen, daß sich der Punkt (x_n, y_n, z_n) im Innern von S befindet. Die Gerade $\mathcal{L}$, die durch die Gleichungen

$$x = x_n + u_n t\,, \quad y = y_n + v_n t\,, \quad z = z_n + w_n t$$

gegeben ist und längs der sich das Teilchen bewegt, möge die Fläche S_i schneiden. Dabei können sich die Eigenschaften des Stoffes ändern.

Man will nun die Länge ϱ_{i_1,i_2} des Intervalls $[T_{i_1}^p, T_{i_2}^q]$ finden, wobei T_i^p der Schnittpunkt der Fläche S_{i_1} mit der Geraden $\mathcal{L}$ und $T_{i_1}^p$ der Schnittpunkt der Fläche S_{i_2} mit der Geraden $\mathcal{L}$ ist (i_2 wird so vorausgesetzt, daß die Gerade $\mathcal{L}$ zwischen T^p und T^q keinen Schnittpunkt mit irgendeiner Fläche hat). Der Größe ϱ_{i_1,i_2} wird der Index desjenigen Stoffes zugeordnet, der zum Intervall $[T_{i_1}^p, T_{i_2}^q]$ gehört. Die Bedeutung der Indizes p und q wird später klar.

Die formulierte Aufgabe werden wir die allgemeine nennen. Zu ihrer Lösung gehen wir zunächst auf den Fall ein, daß es innerhalb von S eine einzige Fläche S_{i_0} gibt, die durch die Gleichungen

$$f_\lambda^{(i_0)}(x, y, z) = 0\,, \qquad \lambda = 0, 1, \ldots, \Lambda\,,$$

und Ungleichungen

$$g_{\mu,\lambda}^{(i_0)}(x, y, z) \geqq 0\,, \qquad \mu = 0, 1, \ldots, M_\lambda\,,$$

beschrieben wird. Um den Schnittpunkt der Geraden $\mathcal{L}$ mit S_{i_0} zu finden, muß man die Gleichung

$$f_\lambda^{(i_0)}(x_n + tu_n, y_n + tv_n, z_n + tw_n) = 0 \tag{3.1.6}$$

lösen und diejenigen reellen Lösungen t nehmen, für die auch die Ungleichungen

$$g_{\mu,\lambda}^{(i_0)}(x_n + tu_n, y_n + tv_n, z_n + tw_n) \geqq 0\,, \qquad \mu = 0, 1, \ldots, M_\lambda\,,$$

erfüllt sind. Die erhaltenen Werte $t_{i_0}^{(p)}$ sind die Koordinaten der Schnittpunkte der Geraden $\mathcal{L}$ mit S_{i_0} vom Punkt (x_n, y_n, z_n) aus gezählt. Anders ausgedrückt ist die Größe $t_{i_0}^{(p)}$, $p = 1, 2, \ldots, N_{i_0}^{(p)}$, die Entfernung vom Punkt (x_n, y_n, z_n) zu einem Schnittpunkt von $\mathcal{L}$ mit S_{i_0}. Ist $t_{i_0}^{(p)} \geqq 0$, so befindet sich der entsprechende Schnittpunkt in der Bewegungsrichtung des Teilchens.

Weiter nehmen wir an:

a) Berührt die Fläche S_{i_0} die Gerade und ist ein gewisser t-Wert mehrfache Wurzel der Gleichung (3.1.6) ist, so zählen wir diesen t-Wert zweimal in der Liste der Werte $t_{i_0}^{(p)}$. Dann ist also N_{i_0} immer eine gerade Zahl.

b) Den Fall, daß ein Teil der Geraden $\mathcal{L}$ der Fläche S_i angehört, schließen wir von der Betrachtung aus. Dazu genügt es in der Regel, eine der Größen $x_n, y_n, z_n, u_n, v_n, w_n$ um einen kleinen Betrag abzuändern. Unter diesen Annahmen ist die Aufgabe, den Intervallen der Geraden $\mathcal{L}$, die innerhalb und außerhalb von S_{i_0} liegen, verschiedene Indizes zuzuordnen, immer lösbar.

Wir nehmen nun an, daß ein Programm vorhanden ist, das bei gegebener Fläche S_{i_0} die Größen $t_{i_0}^{(p)}$, $p = 1, 2, \ldots, N_{i_0}$, berechnet und ihnen irgendeinen Index n_{i_0} zuordnet. Wir bezeichnen dieses Programm mit $\pi(i_0)$. Weiter wird gezeigt, wie aus diesem Programm $\pi(i)$ ein Programm gebildet werden kann, das auch die allgemeine Aufgabe löst.

Wir ordnen die Programme $\pi(i)$ in einer beliebigen aber festen Reihenfolge an (z. B. in der Reihenfolge der Numerierung der Fläche S_i beginnend mit $S = S_0$). Wir nehmen an, daß die als Ergebnis berechneten Werte $t_i^{(p)}$ ein eindimensionales Feld im Speicher der Maschine bilden. Gleichzeitig wird ein anderes Feld gebildet, das Feld der Indizes n_i, so daß jedem $t_i^{(p)}$ ein der Fläche S_i zugeordnetes n_i entspricht.

Der folgende Rechenschritt besteht darin, daß das Feld der Werte $t_i^{(p)}$ der Größe nach geordnet wird. Die entsprechende Permutation wird gleichzeitig auch bei den Indizes vorgenommen, so daß die Beziehung zwischen den $t_i^{(p)}$ und n_i erhalten bleibt. Nach dieser Änderung der Indizierung werden die Elemente der erhaltenen Felder mit $t[j]$ und $n[j]$, $j = 1, 2, \ldots, m$, bezeichnet, wobei $t[m]$ der maximale Wert der $t[j]$ ist. Die uns interessierenden Entfernungen ϱ_{i_1, i_2} werden jetzt mit ϱ_j bezeichnet. Sie sind gleich der Differenz zwischen zwei benachbarten $t[j]$. Die Prozedur zur Berechnung der den Größen ϱ_j entsprechenden Indizes ist nicht so trivial, wenn man berücksichtigt, daß die Flächen S_i ineinander liegen können. Man muß die Felder $t[j]$ und $n[j]$ in der Reihenfolge fallender j-Werte bis zum ersten negativen $t[j]$ durchsehen und sich dabei die Indizes merken, die allen Flächen entsprechen, innerhalb derer sich das Teilchen im gegebenen Moment befindet.

Wir schreiben den Algorithmus zur Berechnung der Entfernungen ϱ_j und der ihnen entsprechenden Indizes als Prozedur in ALGOL 60. Als Parameter der Prozedur erscheinen die Dimension m der Felder $n[j]$ und $t[j]$ und die Dimension R des Feldes $r[i]$, in dem die Indizes der Flächen gespeichert werden, innerhalb derer sich im gegebenen Moment das Teilchen befindet. Wir nehmen an, daß der Fläche $S = S_0$ der Index 0 zugeordnet wird, um die Allgemeinheit des Algorithmus zu wahren.

```
procedure   Geometry (m, R);
        integer  m, R;
        begin integer i, j; real array t[1:m];
                integer array n[1:m], r[1:R];
                r[1] := 0;
           for j := m, j - 1 while t[j] > 0 do
           begin if t[j - 1] ≦ 0 then t[j - 1] := 0;
              t[j] := t[j] - t[j - 1];
           if r[1] = 0 ∧ n[j] = 0 then goto A;
```

```
        if n[j] = r[1] then
        begin for i := 1 step 1 until R - 1 do r[i] := r[i + 1]
        end else
        begin for i := 1 step 1 until R - 1 do
          r[R - i] := r[R - i - 1]; r[1] := n[j]
        end;
        n[j] := r[1];
    A : end end Geometry;
```

Nach Abarbeitung des Programms befinden sich die Entfernungen ϱ_j in den Speicherplätzen $t[j]$ und die ihnen entsprechenden Indizes in den Plätzen $n[j]$. Die ϱ_j sind von $t[m]$ aus in Richtung fallender j bis zum ersten negativen Wert $t[j]$ angeordnet. Einen solchen negativen Wert gibt es sicher, da sich der Punkt x_n, y_n, z_n im Innern von S befindet.

§ 2. Modellierung von Aufgaben der Massenbedienung

Wie schon in § 1 beschränken wir uns auf genügend einfache Beispiele. Aufgaben der Massenbedienung werden in der Sprache der Wahrscheinlichkeitstheorie formuliert. Daher enthält die mathematische Beschreibung einer solchen Aufgabe bereits in impliziter Form die Methode zu ihrer Modellierung. Der analytische Apparat der Theorie der Massenbedienung erlaubt bei weitem nicht die Lösung aller Aufgaben von praktischem Interesse. Daher hat die Entwicklung von Modellierungsmethoden für solche Aufgaben eine große Bedeutung.

Ein Bedienungssystem wird folgendermaßen charakterisiert:

a) Durch den Forderungsstrom. Die Natur dieser Forderungen wird nicht berücksichtigt. Der Strom wird durch die gemeinsame Verteilung der Zeitintervalle zwischen dem Eintreffen zweier aufeinanderfolgender Forderungen beschrieben.

b) Durch eine gewisse Menge von Bedienungsgeräten. Jedes Gerät wird durch die Bedienungszeit charakterisiert. Diese Zeit ist eine Zufallsgröße und wird durch ihre Verteilungsfunktion charakterisiert. Forderungen, die zu einem Zeitpunkt eintreffen, zu dem die Geräte besetzt sind, bilden eine Warteschlange.

c) Durch die Organisationsform beim Warten und Bedienen. Diese kann auch durch eine gewisse Wahrscheinlichkeitsverteilung charakterisiert sein.

Eine Menge von Bedienungssystemen bildet ein Netz, wenn die aus einem System austretenden Forderungen in ein anderes System eintreten. Als einfachstes, ausführlich untersuchtes Beispiel für ein Bedienungssystem kann ein System mit einem Bedienungsgerät betrachtet werden, für das die Zeitintervalle τ_i zwischen zwei unmittelbar hintereinander eintreffenden Forderungen unabhängig und mit dem Parameter λ exponentialverteilt sind, die Bedienungszeit ebenfalls

exponentialverteilt ist mit dem Parameter μ und die Forderungen eine Warteschlange in der Reihenfolge des Eintreffens bilden. Das System wird durch diejenigen Wahrscheinlichkeiten $p_n(t)$ charakterisiert, daß sich zur Zeit t im System (d. h. in der Warteschlange und in der Bedienung) n Forderungen befinden. Ist $\lambda/\mu < 1$, so streben die Wahrscheinlichkeiten $p_n(t)$ für $t \to \infty$ gegen einen Grenzwert, der nicht von t abhängt (stationäres Verhalten). In dem beschriebenen Fall lösen die analytischen Methoden vollständig die Aufgabe. Im stationären Fall kann man eine sehr einfache Formel für die p_n erhalten. Ist aber $t = t_0$ nicht sehr groß und soll $p_n(t_0)$ vor Einstellung des stationären Verhaltens bestimmt werden, so sind die analytischen Formeln bereits recht kompliziert, und eine Modellierung scheint sogar bei dieser einfachen Aufgabe vorteilhaft. Man erzeugt bei der Modellierung Realisierungen der exponentialverteilten Zufallsgrößen und berechnet dann die Anzahl der Forderungen im System zum Zeitpunkt t_0. Indem man diese Modellierung N-mal mit unabhängigen Folgen von Zufallszahlen durchführt, kann man die Wahrscheinlichkeit $p_n(t_0)$ abschätzen und ein Histogramm für die Verteilung der Anzahl der Forderungen im Moment t_0 anfertigen. Weil n die Werte von Null bis Unendlich annimmt, ist es praktisch unmöglich, auf diese Weise alle $p_n(t_0)$ (oder auch nur genügend viele) mit befriedigender Genauigkeit zu erhalten. Die relative Genauigkeit der $p_n(t_0)$ ist, wie meistens, um so größer, je größer $p_n(t_0)$ ist.

In der Regel berechnet man bei der Modellierung eines Bedienungssystems nicht $p_n(t)$, sondern andere Charakteristiken, wie etwa die mittlere Wartezeit in der Schlange auf dem Zeitintervall $[0, t_0]$ usw. In diesen Fällen betrachtet man N unabhängige Abläufe des Systems, jeweils vom Zeitpunkt Null bis zum Zeitpunkt t_0. Sollen Charakteristiken bestimmt werden, die mit dem stationären Verhalten verknüpft sind, so muß man den Ablauf genügend lange verfolgen. Im Grenzfall $t \to \infty$ eines einzelnen Ablaufs werden die Zufallsgrößen, deren Mittelwerte berechnet werden, abhängig, was bei der Fehlerabschätzung der Modellierung und bei der Herleitung von Schätzungen der uns interessierenden Mittelwerte berücksichtigt werden muß (Aufgabe B).

Wir gehen auf diese Frage noch etwas ausführlicher ein. Gegeben ist ein System mit k Geräten und Poissonschem Forderungsstrom (d. h., die Zeit zwischen dem Eintreffen zweier aufeinanderfolgender Forderungen und die Bedienungszeit an jedem Gerät sind exponentialverteilt). Bezüglich der Organisation der Bedienung wird angenommen, daß jedes freie Gerät mit der gleichen Wahrscheinlichkeit die nächstfolgende Forderung bedient (wenn es einige freie Geräte im Moment des Eintreffens einer Forderung gibt). Die Forderungen bilden eine Warteschlange in der Reihenfolge ihres Eintreffens. Ein typisches Verhalten, wie sich die Zahl der Forderungen im System ändert, ist in Abb. 3.4 dargestellt.

Uns interessiert nun die Wahrscheinlichkeit dafür, daß Forderungen nicht sofort bedient werden können (Wahrscheinlichkeit des Blockiertseins). In Abb. 3.4 sind die Forderungen über der punktierten Linie blockiert. Man kann zeigen (Cox und Smith [1]), daß unter den bezüglich des Systems getroffenen Voraussetzungen das ganze Zeitintervall in unabhängige Etappen zerlegt werden kann, jeweils vom Moment des Austritts des Systems aus einem Zustand mit dem Index n bis zum Eintritt in einen analogen Zustand. Die Anzahlen der blockierten Forderungen auf einer bestimmten Etappe bilden unabhängige Zufallsgrößen.

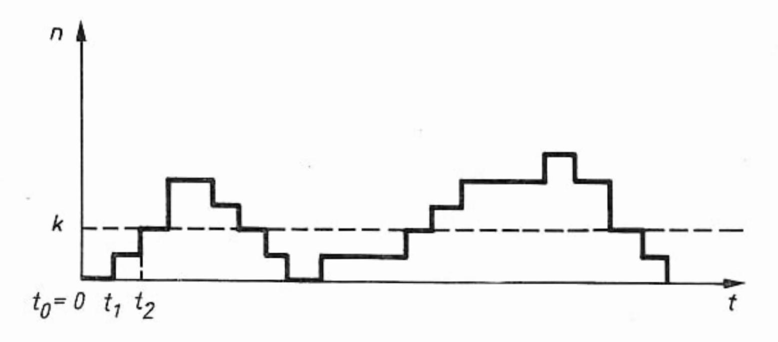


Abb. 3.4

Bezeichnet N_i die Anzahl der auf der Etappe mit dem Index i eintreffenden Forderungen und n_i die Anzahl der blockierten Forderungen auf derselben Etappe, so ergibt sich für die Häufigkeit des Blockierens auf der i-ten Etappe $\hat{p}_i = n_i/N_i$, $i = 1, \dots, M$, wobei M die Anzahl der gewählten Etappen ist. Für den einfachsten Fall nur eines Bedienungsgerätes kann man zeigen, daß für $M \to \infty$ die Schätzung $\tilde{p} = M^{-1} \sum_{i=1}^{M} \hat{p}_i$ verzerrt ist, obwohl die Schätzung (vgl. Kap. I)

$$\bar{p} = \frac{\sum_{i=1}^{M} n_i}{\sum_{i=1}^{M} N_i} \tag{3.2.1}$$

asymptotisch erwartungstreu (unverzerrt) ist. Wird der Forderungsstrom eines Systems durch den Parameter λ charakterisiert und ist die Verteilungsfunktion der Bedienungszeit $1 - \exp(-x)$, so beträgt die Wahrscheinlichkeit des Blockiertseins $\lambda/(1 + \lambda)$. Dagegen ist $\mathsf{M}\hat{p}_i = 1 - \lambda^{-1} \log(1 + \lambda)$. Die entsprechenden analytischen Berechnungen werden beispielsweise in der Arbeit von Kabak [1] durchgeführt. Eben dort wird auch ein Näherungswert für die Dispersion der Schätzung (3.2.1) erhalten. Die Benutzung eines solchen Wertes kann notwendig werden, wenn das Zeitintervall im voraus festgelegt wird und man die Mittelwerte verschiedener Größen auf diesem Intervall abschätzen soll. Dann sind sowohl $n = \sum_{i=1}^{M} n_i$ wie auch $N = \sum_{i=1}^{M} N_i$ Zufallsgrößen. Durch Taylorentwicklung des Bruches n/N in der Umgebung des Mittelwertes erhalten wir

$$\mathsf{D}\left(\frac{n}{N}\right) \approx \frac{(\mathsf{M}(n))^2\, \mathsf{D}(N)}{(\mathsf{M}(N))^4} + \frac{\mathsf{D}(n)}{(\mathsf{M}(N))^2} - \frac{2\mathsf{M}(n) \operatorname{cov}(n, N)}{(\mathsf{M}(n))^2}. \tag{3.2.2}$$

Diesen Ausdruck kann man verwenden, wenn die höheren Momente der Zufallsgrößen n und N klein sind. Die Schätzung der Größen, die in (3.2.2) eingehen, läßt sich im Verlauf der Rechnung recht einfach durchführen. Die Benutzung von Überlegungen über die Unabhängigkeit der Etappen kann also bei der Mo-

dellierung zu falschen Ergebnissen führen. Eigentlich wird die Wahrscheinlichkeit des Blockiertseins in der Zeit $[0, \infty]$ durch den Erwartungswert der Zufallsgröße

$$\xi_i = \begin{cases} 1, & \text{wenn die } i\text{-te Forderung nicht sofort bedient wird,} \\ 0 & \text{sonst} \end{cases}$$

bestimmt. Sind N Forderungen eingetroffen, so erhält man eine erwartungstreue und konsistente Schätzung für die Wahrscheinlichkeit des Blockiertseins in der Form

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i \,. \tag{3.2.3}$$

Das stimmt mit der Schätzung (3.2.1) überein. Hierbei sind die ξ_i abhängige Zufallsgrößen.

Die Frage, wie eine effektive numerische Prozedur für die Schätzung der Dispersion von (3.2.1) hergeleitet werden kann, behandeln wir in Kapitel VI.

Wir gehen jetzt auf einen Algorithmus zur Modellierung von solchen Bedienungssystemen ein, die gewöhnlich in der Bedienungstheorie betrachtet werden (Klimow [1]). Als typisches Beispiel kann das folgende dienen. Die Momente des Eintreffens von Forderungen sind $t_1, t_2, \ldots$ Die Größen $\tau_n = t_n - t_{n-1}$, $n = 1, 2, \ldots, t_0 = 0$, werden als unabhängig und identisch mit der Verteilungsfunktion $A(\tau_n) = A(\tau)$ verteilt vorausgesetzt. Sind im Moment des Eintreffens l Geräte frei, so kann die Forderung ein beliebiges davon bevorzugen (etwa das Gerät mit der kleinsten mittleren Bedienungszeit). Wir werden annehmen, daß die Bedienungszeiten z für jede Forderung und jedes Gerät unabhängig identisch mit der Verteilungsfunktion $B(z)$ verteilt sind. Für $l > 0$ soll die Forderung sofort und mit gleicher Wahrscheinlichkeit eines der Geräte besetzen. Für $l = 0$ reiht sich die Forderung in die Warteschlange ein. Man kann verschiedene Regeln für die Bildung einer solchen Schlange betrachten. Einige Forderungen können in der Bedienung eine Priorität besitzen. Man kann auch den Fall betrachten, daß gewisse Forderungen die Bedienung von früher gekommenen Forderungen unterbrechen und das Gerät besetzen. Für unser Beispiel nehmen wir an, daß die Forderungen eine Warteschlange in der Reihenfolge ihres Eintreffens bilden. Die bedienten Forderungen verlassen das System und bilden dabei den Ausgangsstrom.

Die Modellierung beruht auf der Erzeugung von Realisierungen der Zufallsgrößen mit den Verteilungsfunktionen $A(\tau)$ und $B(z)$. Wir erhalten nacheinander Momente, in denen Forderungen eintreffen ($t_1 = \tau_1, t_2 = t_1 + \tau_2, \ldots, t_n = t_{n-1} + \tau_n$) und die Bedienungszeiten $z_1, z_2, \ldots$. Mit u_n bezeichnen wir die Wartezeit der n-ten Forderung bis zum Beginn der Bedienung. Dann ist $t_n + u_n$ der Moment, in dem die Bedienung beginnt, und $t_n + u_n + z_n$ der Moment, in dem sie beendet ist. Wir bezeichnen mit $t_n + w_{n,i}$, $i = 1, \ldots, k$, den Moment, in dem das Gerät mit dem Index i die Bedienung von Forderungen, die zu den $n - 1$ zuerst im System eingetroffenen gehörten, beendet hat. Mit $\overline{w}_{n,i}$ bezeichnen wir die Zeitdauer von t_n bis zur Beendigung der Bedienung aller vor dem Moment t_n gekommenen Forderungen auf dem i-ten Gerät. Dann ist $u_n = \min \{\overline{w}_{n,1}, \ldots, \overline{w}_{n,k}\}$.

Der Vektor $\overline{\boldsymbol{w}}_n = (\overline{w}_{n,1}, \ldots, \overline{w}_{n,k})$ beschreibt den Zustand, der sich im Ablauf der Bedienung auf jedem der Geräte herausbildet. Zu Beginn ist $\overline{\boldsymbol{w}}_1$ bekannt

(gewöhnlich ist es der Nullvektor). Während der weiteren Arbeit möge nun das Gerät mit dem Index i_0 die n-te Forderung bedienen, also $\overline{w}_{n,i_0} = \min\{\overline{w}_{n,1}, \ldots, \overline{w}_{n,k}\}$. Das wird berücksichtigt, indem man zur i_0-ten Komponente die Größe z_n addiert. Die nächste Änderung erfolgt nach dem Zeitabschnitt τ_{n+1}, wenn die folgende Forderung eintritt. Wir erhalten $\overline{w}_{n+1}$, indem wir von allen Komponenten von $\overline{w}_n'$ die Größe τ_{n+1} subtrahieren (nachdem zur i_0-ten Komponente z_n addiert wurde). Negative Komponenten werden durch Null ersetzt. Eine der minimalen Komponenten von $\overline{w}_{n+1}$ (es können einige Komponenten gleichen Wert haben) wird als Größe u_{n+1} ausgewählt. Indem man so die Arbeit des Systems verfolgt, kann man verschiedene uns interessierende Charakteristiken bestimmen. Sind N Forderungen eingetroffen, so kann der Wert $N^{-1} \sum_{n=1}^{N} u_n$ als Schätzung für die mittlere Wartezeit einer Forderung bis zum Bedienungsbeginn dienen. Analog konstruiert man Schätzungen für andere Größen (Funktionale). Gewöhnlich interessiert man sich für solche Größen, wie die mittlere Anzahl der auf Bedienung wartenden Forderungen, die mittlere Anzahl der freien Geräte, die mittlere Verweildauer einer Forderung im System usw.

Außer der Exponentialverteilung mit konstanten oder zeitabhängigen Parametern begegnet man in der Bedienungstheorie besonders oft der Erlangverteilung (Gammaverteilung) und der Hyperexponentialverteilung. Die erste wird durch die Parameter k und λ charakterisiert (λ kann von der Zeit abhängen). Die entsprechende Wahrscheinlichkeitsdichte ist $p_k(x, \lambda) = \lambda^k x^{k-1} e^{-\lambda x}/(k-1)!$. Es ist leicht zu sehen, daß die Summe unabhängiger identisch exponentialverteilter Größen (also mit der Dichte $\lambda e^{-\lambda x}$) eine solche Dichte besitzt. Die zweite wird durch die Dichte

$$p(x) = \sum_{i=1}^{k} p_i \lambda_i \exp(-\lambda_i x) \quad \text{mit} \quad \sum_{i=1}^{k} p_i = 1\,, \quad p_i > 0\,, \quad x \in [0, \infty)$$

charakterisiert. Ebenso wie die λ_i können auch die p_i von der Zeit abhängen. Für die Modellierung der Hyperexponentialverteilung kann man Kompositionsmethoden gut verwenden.

Gewöhnlich werden mit der Monte-Carlo-Methode kompliziertere Bedienungssysteme als die oben beschriebenen und Bedienungsnetze modelliert. Die Ausgangsverteilungen müssen dabei, wie auch in den früher beschriebenen Aufgaben des Strahlendurchgangs, experimentell hergeleitet und vorher überarbeitet werden. Wenn man etwa die Bedienung von Fahrgästen durch eine Autobuslinie modelliert, dann muß man experimentell die Charakteristiken des Fahrgästestroms an jeder Haltestelle bestimmen. Wird die Arbeit eines Hafens modelliert, so muß man vorher die Angaben über den Strom der ankommenden Schiffe und die Verteilung ihrer Entladungszeit (Bedienung) ermitteln. Oft hat die Aufgabe einen ökonomischen Aspekt — man soll feststellen, wie notwendig eine Änderung der Charakteristiken (Anzahl) der Bedienungsgeräte oder der Bedienungsordnung ist, um die mittlere Bedienungszeit zu verkleinern oder, allgemeiner, eine gewisse Kostenfunktion zu minimieren.

Eine wichtige Rolle spielt bei der Modellierung von Bedienungsaufgaben der algorithmische Aspekt. Man soll in einer zweckmäßigen Bezeichnungsweise das

Modell des Systems beschreiben. Hat man es mit einer bestimmten Klasse von Systemen zu tun, so kann ein System aus Blöcken oder Prozeduren in einer algorithmischen Sprache derart gebildet werden, daß schnell Programme für Aufgaben der gegebenen Klasse aufgebaut werden können. Zum Zweck der Modellierung wurden auch spezielle algorithmische Sprachen ausgearbeitet (z. B. SIMSKRIPT (vgl. Markowitz u. a. [1]), SIMULA). Damit gibt es zu dieser Frage eine umfangreiche und leicht zugängliche Literatur.

Wir werden nun eines der einfachsten Beispiele für ein reales System betrachten, in dem eine Modellierung nützlich erscheint.

Beispiel 1. *Bedienung von Werkbänken durch einen Brückenkran.* Brückenkräne stellen ein wichtiges Transportmittel beim Maschinenbau dar. Ihre Arbeit wird folgendermaßen organisiert: In 5—7 m Höhe über der Werkhalle befinden sich einer oder mehrere Brückenkräne. Die Brücke des Kranes, die die Werkhalle überspannt, kann sich nur in Längsrichtung bewegen. Der Lasthaken des Krans bewegt sich nur in Richtung der Brücke, d. h. quer zur Werkhalle. Die gestrichelten Quadrate in Abb. 3.5 sollen Werkbänke darstellen. Diese werden durch den Kran bedient. Der Kran muß zu einer Werkbank gelangen und dort einige Zeit für die Bedienung aufwenden können, wenn das die Produktion erfordert. Wenn der Kran *2* die Werkbank *B* bedient, kann offensichtlich der

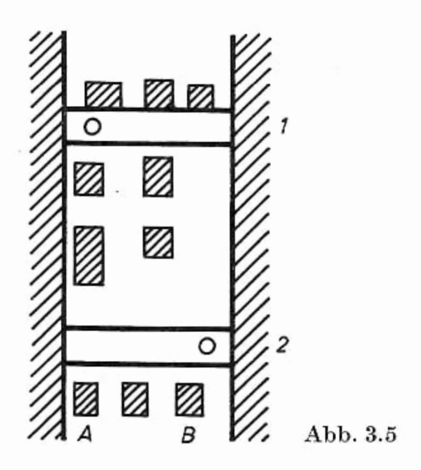


Abb. 3.5

Kran *1* die Werkbank *A* nicht bedienen, d. h., die Kräne können sich gegenseitig stören. Die Werkbänke stellen also Forderungen und die Kräne sind Bedienungsgeräte. Ein Bedienungsgerät ist nicht immer erreichbar, d. h., das Bedienungssystem ist nicht vollständig zugänglich. In der gegebenen Aufgabe interessiert man sich für die Wahl einer geeigneten Bedienungsordnung und für die optimale Anzahl von Kränen, die in der Werkhalle aufzustellen sind. Versuche, eine solche Aufgabe mit „gesundem Menschenverstand" zu lösen, führen gewöhnlich zu groben Fehleinschätzungen. Die analytische Lösung einer Aufgabe der Bedienungstheorie für unvollständig zugängliche Systeme bietet sogar bei einem Poissonschen For-

derungsstrom und exponentiell verteilter Bedienungszeit erhebliche Schwierigkeiten. Das Modell dieses Vorgangs ist dagegen recht einfach. Wir legen die Bedienungsordnung fest. Es wird z. B. die nächstliegende Werkbank bedient, von der eine Forderung gestellt wurde, oder der gegebene Kran bedient nur bestimmte Werkbänke. Wir geben die Anzahl der Kräne und die Koordinaten der Werkbänke vor. Ist der Forderungsstrom von jeder Werkbank Poissonsch, so ergibt auch die Summe einen Poissonschen Strom. (Im einfachsten Fall stammt die nächste Forderung gleichwahrscheinlich von jeder Werkbank.) Das Schema des Algorithmus ist recht einfach, so daß ihn der Leser leicht selbst aufstellen kann. Soll die Aufgabe nach der optimalen Anzahl der Kräne gelöst werden, so muß ein Optimalitätskriterium angegeben werden, das die Kosten einer stillstehenden Werkbank und die Arbeitskosten eines Kranes berücksichtigt. Ist k die Anzahl der Werkbänke und l die Anzahl der Kräne, so kann man

$$\varphi(l) = \frac{kz_1 + lz_2}{I}$$

als ein solches Kriterium ansehen. Dabei sind z_1 die Ausgaben, die mit dem Stillstand einer Werkbank verbunden sind, z_2 die Ausgaben, die mit dem Betrieb eines Kranes verbunden sind, und I ist die mittlere Anzahl der arbeitenden Werkbänke. Die Besonderheit bei der Berechnung besteht darin, daß die Erfüllung einer Ungleichung $\varphi(l_1) > \varphi(l_2)$ geprüft werden muß, d. h., man muß ein genügend enges Konfidenzintervall für die Schätzung von $\varphi(l)$ gewährleisten. Hier haben Fragen der Dispersionsschätzung erstrangige Bedeutung.

Methoden der Bedienungstheorie und die entsprechenden Modellierungsmethoden finden oft Anwendung in Gebieten, die auf den ersten Blick von der eigentlichen Bedienungstheorie weit entfernt liegen. In dieser Beziehung ist folgendes Beispiel interessant:

Beispiel 2. *Anwendung auf Aufgaben der seismischen Erkundung.* In der Arbeit von Solotuchina u. a. [1] wird ein stochastisches Modell für den Durchgang von Explosionswellen durch Medien aus mehreren Schichten beschrieben. Das Modell hat eine wichtige Bedeutung für die seismische Erkundung, bei der aus dem Charakter der reflektierten Wellen die Struktur von Lagerstätten beurteilt werden sollen.

Auf der Erdoberfläche sollen längs einer Geraden Detonationsquellen und Empfänger angeordnet sein. (Die Erdoberfläche wird als eben betrachtet.) Durch die genannte Gerade wird eine Ebene senkrecht zur Erdoberfläche gelegt und der so erhaltene Schnitt betrachtet.

Die z-Achse ist in das Erdinnere gerichtet, Q sei eine Detonationsquelle, und mit E bezeichnen wir Punkte, in denen sich Empfänger befinden. Man nimmt an, daß die Schichten, die die Wellen reflektieren, senkrecht zur Schnittebene und ihre Grenzen parallel zur x-Achse liegen. Medien solcher Art trifft man in der geologischen Praxis in Form von Sedimentgesteinen.

Besonderes Interesse beanspruchen gewöhnlich ein oder zwei Schichten genügend großer Ausdehnung (sie sind nicht in Abb. 3.6 dargestellt). Die an ihnen reflektierten Wellen liefern nützliche Signale. Außerdem gibt es eine Reihe anderer Schich-

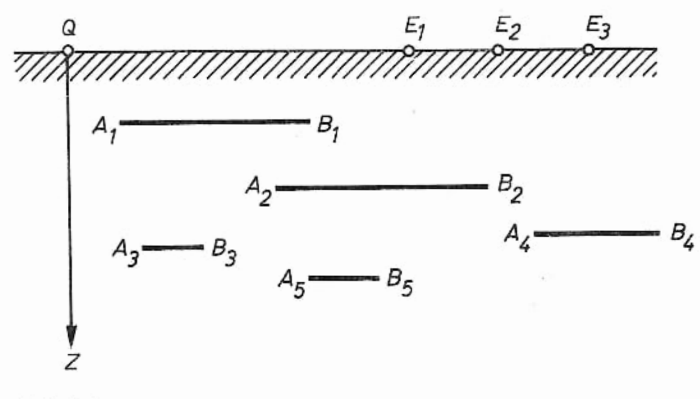


Abb. 3.6

ten, die als Störungsquellen dienen. Die Projektionen ihrer Grenzen auf die Schnittebene sind durch die Intervalle A_kB_k dargestellt. Die Aufklärung des Störungscharakters hat große Bedeutung für die Aussonderung der nützlichen Signale.

Wir werden ein Intervall A_kB_k charakterisieren durch die Länge l_k, die Tiefe der Lagerstätte z_k und die Projektion x_k des Punktes A_k auf die x-Achse und nehmen an:

1. Die Werte x_k bilden einen zufälligen Strom (von Forderungen), und $S(x)$ ist gleich der Anzahl der x_k auf dem Intervall $[0, x]$.

2. Die Zufallsgrößen l_k sind identisch verteilt mit der Verteilungsfunktion $F_l(x) = \mathsf{P}(l \leqq x)$.

3. Die Zufallsgrößen z_k, $0 \leqq z_k \leqq H$, haben die Verteilungsfunktion $G(x) = \mathsf{P}(z_k \leqq x)$.

4. Die Reflexion geht nach den Regeln der geometrischen Optik vor sich. Die reflektierte Welle beschreibt man durch eine bekannte Zeitfunktion, die durch die Formel

$$Af(t-\tau) = \begin{cases} 0 & \text{für} \quad t \leqq \tau\,, \\ A\bar{f}\,(t-\tau) & \text{für} \quad t > \tau \end{cases}$$

gegeben wird. Dabei ist τ die Zeit bis zum Eintreffen der Welle und A die zufällige Amplitude mit gegebener Verteilungsfunktion.

Es wird auch vorausgesetzt, daß die Größen z_k, l_k, A in der Gesamtheit und von x_k unabhängig sind und das zweite Moment der entsprechenden Verteilungsfunktion endlich ist.

Nimmt man an, daß die x_k einen Poissonschen Strom bilden, die l_k exponentialverteilt und die z_k auf $[0, H]$ gleichverteilt sind, so scheint die betrachtete Aufgabe den traditionellen Aufgaben der Bedienungstheorie sehr ähnlich zu sein, wobei die reflektierten Wellen die Rolle der Bedienungsgeräte übernehmen.

Die praktische Berechnung ist auf der Grundlage der analytischen Resultate, die Solotuchina u. a. [1] erhalten haben, allerdings nur schwer durchführbar.

Daher erscheint die unmittelbare Modellierung des Verlaufs der Wellen bequemer. Das Schema der Modellierung besteht in folgendem: Entsprechend den gegebenen Verteilungsfunktionen wird eine Strukturkarte der Erdrinde aufgestellt, d. h., man erhält Realisierungen der Größen

$$x_1, \ldots, x_n; \quad l_1, \ldots, l_n; \quad z_1, \ldots, z_n .$$

Weiter wird die Länge des Weges und damit die Zeit der Verzögerung für die an verschiedenen Grenzen reflektierten Wellen bestimmt. Dabei sollen die Ausbreitungsgeschwindigkeit der Wellen und der Brechungskoeffizient gegebene konstante Größen sein. Wir betrachten zunächst die einfache Reflexion der Wellen. Ist CD eine zur z-Achse parallele Linie, die das Intervall QE halbiert (Abb. 3.7), so stellen die punktierten Linien, die durch den Schnittpunkt von CD mit einem Intervall A_kB_k verlaufen, den Weg der entsprechenden Wellen dar. Hat ein Intervall A_kB_k keinen Schnittpunkt mit CD, so gelangt die dort reflektierte Welle nicht zum Punkt E. Weil die reflektierenden Schichten für die Wellen teilweise durchlässig sind, muß man bei gegebenem Brechungskoeffizienten berücksichtigen, daß sich nach der Reflexion die Amplitude der Welle auf den q-ten Teil verringert ($0 \leqq q \leqq 1$); die durch die Schicht hindurchgehende Welle setzt ihren Weg fort, aber sie kann offenbar nicht mehr den Punkt E erreichen. Die Schicht A_1B_1, die die punktierten Linien schneidet, kann nur die Intensität der Wellen schwächen.

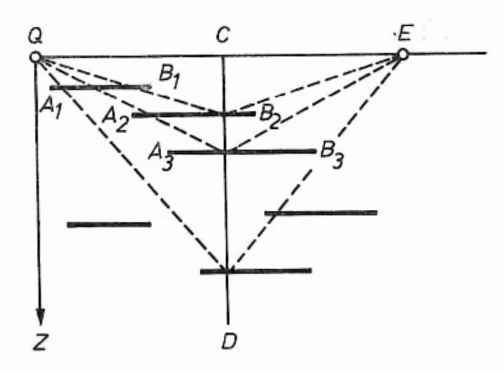


Abb. 3.7

Unter etwas komplizierteren Bedingungen können doppelt oder mehrfach reflektierte Wellen zum Punkt E gelangen (Abb. 3.8).

Obwohl die Reflexionen und Brechungen einen komplizierten verzweigten Prozeß bilden, kann man durch die Bedingung, daß alle Wellen im Punkt E eintreffen müssen, nacheinander und verhältnismäßig einfach die einfach, doppelt usw. reflektierten Wellen aussondern.

Die Überlagerung der mit Verzögerung und veränderten Amplituden im Punkt E ankommenden Wellen ergibt ein gestörtes Signal, das sich zu dem ursprünglichen Signal in E addiert. Mehrfache Realisierung der Struktur der Erdrinde und Bestimmung der Form des Signals erlauben es, allgemeine Eigenschaften der Hindernisse zu bestimmen, die sich aus den Verteilungsfunktionen der Größen x, z, l und A ergeben.

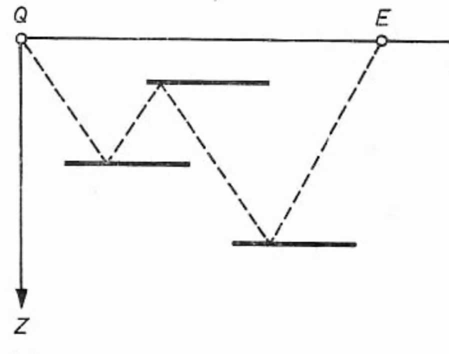


Abb. 3.8

Der Umfang der Rechnungen ist gewöhnlich sehr groß und erfordert die Bestimmung der Abhängigkeit der Ergebnisse von den Parametern der genannten Verteilungen. Wie man in solchen Fällen den Umfang der Rechnungen verkleinern kann, wird in Kapitel V untersucht.

§ 3. Einige andere Beispiele und allgemeine Bemerkungen

Beispiel 1. **Modellierung einer Panzerschlacht.** Das im folgenden beschriebene Beispiel ist der Arbeit von ANDERSSON [1] entnommen. Diese Arbeit wurde im Auftrag der schwedischen Streitkräfte angefertigt. Der Autor nimmt an (wie es scheint, nicht ohne Grund), daß auf dem Territorium Schwedens eine Schlacht unwahrscheinlich ist, an der mehr als 10 bis 15 Panzer von jeder Seite teilnehmen.

Es wird angenommen, daß infolge der hohen Schußzahl je Zeiteinheit und der großen Wahrscheinlichkeit, auf Entfernungen kleiner als 1,5 km zu treffen, das Gefecht sehr schnell abläuft. Dann kann die Bewegung der Panzer während der Schlacht vernachlässigt werden. Daher wird ein Modell betrachtet, in dem die Panzer von festen Positionen aus schießen und eine Bewegung nur am Anfang zulässig ist.

Das ganze Zeitintervall wird in Intervalle gleicher Länge geteilt. Auf jedem Teilintervall wird das „Schlachtfeld" vollständig durchgemustert, und man berücksichtigt alle Änderungen, die durch die Wirksamkeit einzelner kämpfender Einheiten (also durch Panzer) des Modells hervorgerufen werden. Dabei wird angenommen, daß jede Einheit des Modells sich in einem der folgenden Zustände befinden kann:

1. unverletzt;
2. kann nur noch schießen;
3. kann sich nur noch bewegen (Zwischenzustand);
4. ist aus der Front ausgeschieden.

In einem Zwischenzustand befindet sich eine getroffene, aber noch nicht aus der Front ausgeschiedene Einheit. Nach einiger Zeit kann sie in einen der drei vorherigen Zustände übergehen. Außer diesen Zuständen hat jede Einheit noch

andere Eigenschaften. Einige von ihnen sind konstant, wie z. B. die Geschwindigkeit, andere können sich während der Schlacht ändern, z. B. der Typ der Waffen, die im gegebenen Moment benutzt werden, die Schußgeschwindigkeit und einige spezielle Charakteristiken (z. B. die Übergangszeit aus dem Zwischenzustand).

Die Aktivität der Seite A gegenüber der Seite B wird durch die Aktivitätsmatrix $M_{AB} = (\|(m_{AB}(i, j))\|)$ beschrieben. Ihre Elemente können die ganzzahligen Werte von 0 bis 4 annehmen. Bezeichnet A_i die i-te Einheit der Seite A und B_j die j-te Einheit der Seite B, so hat $m_{AB}(i, j)$ die folgenden Werte:

$m_{AB}(i, j) =$	0	1	2	3	4
Hat A_i die Einheit B_j entdeckt?	nein	ja	ja	ja	ja
Hat A_i B_j als Ziel gewählt?	nein	nein	ja	ja	ja
Hat A_i wenigstens einen Schuß auf B_j abgegeben?	nein	nein	nein	ja	ja
Bewegt sich im gegebenen Moment ein Schuß von A_i nach B_j?	nein	nein	nein	nein	ja

So bedeutet etwa $m_{AB}(i, j) = 3$, daß A_i den folgenden Schuß auf B_j vorbereitet.

Die Anfangsphase eines Gefechts ist gewöhnlich sehr wichtig für deren Ausgang, insbesondere ist es wichtig, wie schnell die Seiten sich gegenseitig entdecken. Auf einem offenen Platz vollzieht sich dieses Entdecken fast unmittelbar, hinreichend viele Panzer nähern sich bis auf Sichtweite. Unter anderen Verhältnissen kann die Wahrscheinlichkeit einer Entdeckung von vielen Faktoren abhängen. Steht etwa ein Panzer in fester Position getarnt, so ist es schwer, ihn zu entdecken, bevor er nicht schießt. Die Schaffung eines adäquaten Modells für das Entdecken erfordert gründliche Untersuchungen physikalischer und psychologischer Art. Im betrachteten Modell wird die Entdeckungswahrscheinlichkeit p_{ij} im Zeitintervall $(t - \Delta t, t)$ eingeführt unter der Bedingung, daß A_i vorher (also in $(0, t - \Delta t)$) B_j noch nicht entdeckt hat

$$p_{ij}(t) = \frac{g_{i,j}(t)\left(h_i(t) + S_{j,k}(t)\, K_j\right) \Delta t}{d_{i,j}}.$$

Hierbei ist

$$g_{i,j}(t) = \begin{cases} 0 & \text{für} \quad t < \tau_{i,j}, \\ 1 & \text{für} \quad t \geqq \tau_{i,j}; \end{cases}$$

$\tau_{i,j}$ ist der Moment, in dem die direkte Sichtbarkeit zwischen A_i und B_j möglich wurde; $h_i(t)$ ist die Beobachtungsaktivitätsfunktion von A_i. Diese Funktion ist gleich einer Konstanten bis zum und nach dem Moment $t = t_2$, in dem sie sprunghaft wächst. t_2 ist dabei der Moment, in dem das Erscheinen feindlicher Panzer bekannt wird, und bildet sich aus der Zeit t_1 der ersten Entdeckung des Gegners durch die Seite A und einer (zufälligen) Benachrichtigungszeit. $S_{j,k}(t)$ ist eine Funktion, die den Effekt der Entdeckung im Moment t_k des k-ten Schusses von B_j charakterisiert. Abb. 3.9 entspricht dem Fall, daß im Moment t_3 und t_4 Schüsse abgegeben werden. Die Größe K_j berücksichtigt den Tarnungseffekt und $d_{i,j}$ die Entfernung zwischen A_i und B_j. Die Matrix $(\|(\tau_{i,j})\|)$ wird in der Anfangsphase der Panzerbewegung aufgestellt. Darin werden die örtlichen Bedingungen, das Vor-

handensein von Nebel usw. berücksichtigt. Die Frage, ob die Einheit A_i eine Einheit B_j entdeckt hat, wird also positiv entschieden, wenn $p_{i,j}(t) \leqq \alpha$ wird.

Die Regeln der Zielauswahl waren im gegebenen Modell deterministisch. Es wurden zwei Taktiken angewendet: Eröffnung des Feuers sofort nach Entdeckung des Ziels und Eröffnung des Feuers auf Kommando. Die erste Taktik wurde bei unerwartetem Kontakt angewendet, die zweite bei rechtzeitiger Vorbereitung des Angriffs. Ist das Ziel getroffen oder befindet sich im Übergangszustand, so wurde in Übereinstimmung mit den aufgestellten Regeln das Ziel gewechselt. Dabei gab es in Abhängigkeit von der gegenseitigen Lage der Gegner verschiedene Prioritäten. Der Charakter solcher Prioritäten läßt sich leicht darstellen, aber wir werden darauf nicht näher eingehen.

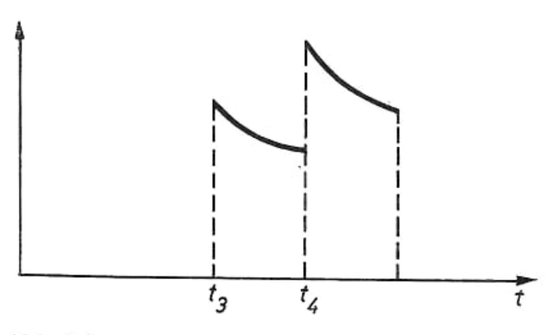


Abb. 3.9

Die Wahrscheinlichkeit, das Ziel getroffen zu haben, wurde als Funktion des Abstands zwischen den Panzern berechnet. Außerdem hängt sie von der Art der Waffe ab, von der Genauigkeit bei der Entfernungsbestimmung und von den Ausmaßen des sichtbaren Zieles. Wenn das Ziel getroffen wurde, erfolgte der Übergang von einem Zustand in einen anderen. Dafür wurde eine Tabelle der Übergangswahrscheinlichkeit in Abhängigkeit von den oben aufgezählten Faktoren für folgende Zustände berechnet:

$$1 \to 1\,, \qquad 2 \to 2\,, \qquad 3 \to 3\,,$$
$$1 \to 2\,, \qquad 2 \to 4\,, \qquad 3 \to 4\,,$$
$$1 \to 3\,,$$
$$1 \to 4\,.$$

Die Genauigkeit bei der Entfernungsbestimmung konnte zwischen zwei Stufen variieren, die Ausmaße des sichtbaren Zieles hingen von der Zeit ab, die seit dem Moment der ersten Sichtbarkeit vergangen war. Dieser Faktor wuchs stufenweise und hatte drei verschiedene Werte. Das Zeitintervall zwischen zwei Schüssen wurde als Summe von zwei zufälligen Größen mit gegebenem Verteilungsgesetz und einem konstanten Summanden (abhängig vom Waffentyp) bestimmt.

Für das beschriebene Modell wurde ein Programm in FORTRAN II für die elektronische Datenverarbeitungsanlage IBM 7090 aufgestellt. Auf jeder Seite

durften höchstens 15 Panzer teilnehmen. Die Eingangsdaten bestanden aus etwa 1000 Zahlen. Die Rechenzeit betrug 18 s bis zum Entscheid eines Gefechts von 15 gegen 9 Panzer, was einer Gefechtsdauer von 100 s mit einer Schrittweite $\Delta t = 1$ s entspricht. Um eine bestimmte Vorstellung von den möglichen Resultaten zu erhalten, mußte das Gefecht etwa 50mal durchgespielt werden.

Das beschriebene Beispiel ist charakteristisch für die Lösung von Aufgaben aus der Spieltheorie. Es ist leicht einzusehen, daß die analytische Beschreibung und Untersuchung ähnlicher Aufgaben praktisch unmöglich ist.

Beispiel 2. *Bestimmen eines Maschinenbelegungsplans für eine technische Einrichtung.* Wir betrachten ein Fließband, das für den Guß von Werkstücken bestimmt ist. In Abb. 3.10 bedeutet F eine Formgebungsmaschine, auf der die Formen für den Guß vorbereitet werden. Die fertigen Formen, die für den Guß einer oder mehrerer Werkstücke bestimmt sind, werden auf dem Band in Pfeilrichtung bewegt. Der Buchstabe S bezeichnet den Schmelzofen. Das Metall aus dem Ofen fließt in die Formen, die danach vom Band heruntergenommen werden und in die weitere Verarbeitung eintreten. Im Verlauf einer Schicht können Gußstücke verschiedener Typen produziert werden. Wir bezeichnen mit m_i ($i = 1, \ldots, n$)

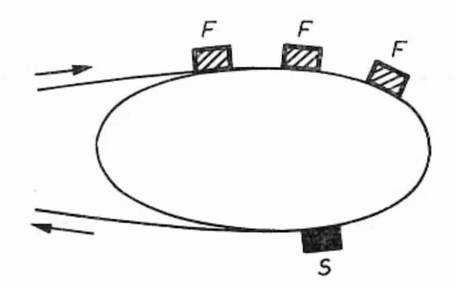


Abb. 3.10

die Anzahl der Stücke vom i-ten Typ, mit p_i das Metallvolumen eines Stücks vom i-ten Typ, mit p die Kapazität des Ofens in einer Schicht und mit t_i die Formungszeit für ein Stück vom i-ten Typ. Es ist notwendig, daß die Kapazität der Öfen und der formgebenden Maschinen in Übereinstimmung stehen. Andernfalls kann es zu einem unproduktiven Stillstand der formgebenden Maschinen kommen, oder aber das Metall reicht nicht. Dabei ist noch zu berücksichtigen, daß ein gewisser Rückstand von Metall (etwa R kg) im Ofen bleiben kann und der Übergang von der Formung eines Werkstücktyps auf einen anderen eine gewisse Zeit $\tilde{t}$ erfordert.

Wir nehmen an, daß die Ungleichungen

$$\sum_{i=1}^{n} p_i m_i \leqq p \quad \text{und} \quad \sum_{i=1}^{n} t_i m_i + n\tau \leqq kT \tag{3.3.1}$$

erfüllt sind, wobei k die Anzahl der formgebenden Maschinen und T deren Arbeitszeit bedeutet. Es soll eine solche Folge von Formgebungen eines Loses von Werkstücken für jede Maschine ausgewählt werden, daß

1. es keinen Stillstand der formgebenden Maschinen gibt;

2. das Metall ausreicht (d. h., es kommt zu keiner Warteschlange von Formen vor dem Ofen);

3. der Rückstand von Metall minimal ist und niemals R übersteigt.

Abb. 3.11 stellt eine Situation dar, die bei der Arbeit nur einer formgebenden Maschine entsteht. Das erste Los enthält ein Stück mit höherem Metallvolumen, und der Metallverbrauch übersteigt das mittlere Niveau. Die folgenden Lose haben dagegen ein niedrigeres Metallvolumen, und wenn sie gegossen werden,

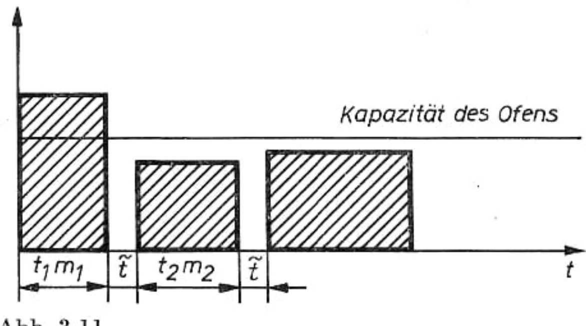


Abb. 3.11

bildet sich ein gewisser Metallvorrat. Gibt es nur eine formgebende Maschine, so sollte man die Arbeit mit Stücken kleinen Metallvolumens beginnen, und wenn sich genügend Metall angesammelt hat, kann man Formen höheren Metallvolumens gießen. Gibt es dagegen mehrere formgebende Maschinen, so kann ein Metallmehrverbrauch, der durch Lose einer formgebenden Maschine hervorgerufen wird, durch kleines Metallvolumen der Lose einer anderen ausgeglichen werden. Bei großer Anzahl von Losen und von formgebenden Maschinen ist die Aufgabe recht kompliziert. Man kann versuchen, sie mit der Monte-Carlo-Methode zu lösen, die im gegebenen Fall auf die Konstruktion einer zufälligen Folge von Losen und auf die Prüfung von deren Zulässigkeit führt. Eine solche zufällige Folge kann folgendermaßen gebildet werden.

Man wählt eine Zufallszahl α_1. Der ganze Teil der Zahl $n\alpha_1$ bezeichnet die Nummer des Loses, das als erstes auf der ersten formgebenden Maschine bearbeitet werden soll. Danach wählt man α_2. Der ganze Teil der Zahl $(n - 1)\,\alpha_2$ wird die Nummer des Loses (unter den verbliebenen Losen), das als erstes für die zweite formgebende Maschine genommen wird, usw. Indem man so eine Reihenfolge der Lose herstellt, erhält man einen gewissen Plan. Schon nach der Wahl des ersten Loses für jede Maschine kann man die Zulässigkeit des Planes beurteilen und die weitere Auswahl verkürzen, wenn die Bedingungen 1.—3. auf einem gewissen Zeitintervall nicht erfüllt sind. Bei solcher Vorgehensweise kann man erfolgreich sein, wenn es in der Menge der angeordneten Lose genügend viele gibt, die unseren Forderungen genügen, d. h., wenn die Wahrscheinlichkeit verhältnismäßig groß ist, in der Menge aller Pläne einen zulässigen Plan zu finden (wobei jeder Plan mit der gleichen Wahrscheinlichkeit gewählt wird).

Es ist leicht zu sehen, daß die gleichwahrscheinliche Auswahl nicht die besten Vorgehensweise darstellt. Gibt es zwei formgebende Maschinen und wurde auf

der ersten ein Los mit großem Metallvolumen gewählt, so sollte man offensichtlich in derselben Zeit auf der zweiten ein Los mit kleinem Metallvolumen wählen. Diese Überlegung wie auch andere Überlegungen solcher Art kann man berücksichtigen, indem man gewissen Loskombinationen eine große Wahrscheinlichkeit zuordnet und die Auswahl der Folgen entsprechend dieser Wahrscheinlichkeitsverteilung organisiert. Die einfachste Methode besteht darin, daß man im Fall von zwei Maschinen auf die erste Maschine vorzugsweise Lose mit großem Metallvolumen und auf die zweite Lose mit kleinem Metallvolumen nimmt. Als Grenzfall erscheint der Fall, daß wir genau wissen, wie ein zulässiger Plan zu konstruieren ist, der dann die Wahrscheinlichkeit 1 bekommt.

Die beschriebene Aufgabe ist nicht in allen Einzelheiten formalisiert. Die Monte-Carlo-Methode kann in diesem Beispiel den Erfolg nicht in allen Fällen garantieren. Gelingt es uns aber, eine befriedigende Lösung zu finden, so ist die Nützlichkeit dieser Methode nicht zu bezweifeln. Besonders wertvoll ist die Möglichkeit, bei der beschriebenen Methode auch intuitive Überlegungen berücksichtigen zu können. Offensichtlich wäre unser Überblick über die Anwendungen der Monte-Carlo-Methode unvollständig gewesen, wenn wir die Beschreibung von Anwendungen dieser Art weggelassen hätten.

Es ist leicht, weitere Beispiele zu finden, in denen die Modellierung von Verteilungen für die Untersuchung realer Erscheinungen ausgenutzt wird. Wir begnügen uns mit der kleinen Anzahl der aufgezählten Beispiele, weil einerseits die unmittelbare Modellierung und die Struktur von Modellen nicht Hauptanliegen des Buches sind und es andererseits eine genügend umfangreiche und gut zugängliche Literatur zu diesen Fragen gibt. So werden eine Reihe von Modellen in den Büchern von Buslenko [1], [2] besprochen. Im Kapitel XVII des ersten Buches von Buslenko ist ein ausgezeichnetes Beispiel für die Modellierung der Arbeit eines Hafens enthalten, das aus dem Artikel von Star und Page [1] entnommen wurde. In der Monographie von Klimow [1] werden ziemlich ausführlich Probleme der Modellierung komplizierter realer Bedienungssysteme erörtert. Der Modellierung von Bedienungssystemen ist auch das Kapitel X der Monographie von Kaufmann und Cruon [1] gewidmet. Zu nennen wäre auch die umfangreiche Monographie von Chorofac [1], die allgemeinen Fragen der Modellierung gewidmet ist, sowie der Sammelband „Methoden der numerischen Physik", in dem sich viele Arbeiten mit der Anwendung der Monte-Carlo-Methode auf physikalische Aufgaben beschäftigen. Diese Liste ist bei weitem nicht vollständig und enthält nur solche Arbeiten, die eine Beziehung zu den aufgezählten Beispielen haben.

Wie schon bemerkt, stellten wir uns in diesem Kapitel nicht das Ziel, die mathematische Untersuchung der zu modellierenden zufälligen Prozesse vorzunehmen, und speziell haben wir betont, daß eine solche Untersuchung in bestimmten Maße die Modellierung selbst ersetzen kann. Ein solches „experimentelles" Vorgehen rückt zwei Fragen in den Vordergrund: die Frage nach vorteilhaften Algorithmen für die Modellierung, die wir im zweiten Kapitel betrachteten, und die Frage nach einer bequemen Beschreibung der Modellstruktur. Teilweise haben wir diese Fragen im vorliegenden Kapitel behandelt. Dabei beschränkten wir uns freilich auf einige Beispiele verschiedener Struktur. Diese beiden Fragen haben oft die Auf-

merksamkeit der Forscher von der wichtigen Aufgabe einer Fehlerabschätzung für die erhaltenen Resultate abgelenkt. Dieser Fehler setzt sich zusammen aus dem Modellfehler und dem statistischen Fehler bei der Modellierung.

Um den statistischen Fehler herauszubekommen, muß man die statistische Reproduzierbarkeit der Resultate mit der notwendigen Genauigkeit erreichen. Für seien Abschätzung ist es besonders wichtig zu wissen, mit welchem Typ von zufälligen Größen wir es zu tun haben. Die Aufgaben des Typs A, B, C und D, die im ersten Kapitel formuliert wurden, erfordern ein unterschiedliches Herangehen an die Fehlerabschätzung, obgleich man sich oft einfach damit begnügt, bei der Wiedergabe der Modelle eine Reihe approximativer Resultate zu erhalten.

Die Abschätzung des Modellfehlers (wenn vom Fehler der Modellierung abgesehen wird) erfordert genau genommen den Vergleich mit dem Experiment. Es kann vorkommen, daß solche Experimente früher nicht durchgeführt wurden oder nicht durchführbar sind, weil sie zu kompliziert und zu kostspielig sind, obwohl es natürlich experimentelle Daten gibt, die für die näherungsweise Beschreibung der elementaren Verteilungen benötigt wurden. Da der resultierende Fehler sich aus den Fehlern der einzelnen Verteilungen zusammensetzt, sollte man in diesen Fällen wenigstens experimentell die Stabilität des Modells in bezug auf die Fehler der Verteilungen überprüfen. Diese Prüfung wird gewöhnlich mit der Monte-Carlo-Methode realisiert, d. h., die einzelnen Verteilungen oder ihre Parameter werden mit gewissen zufälligen Fehlern versehen, worauf dann die Modellierung durchgeführt wird. Die Frage nach der Verteilung der Fehler wird oft intuitiv gelöst und seltener auf der Grundlage einer Analyse der experimentellen Daten.

Nicht selten sind die Daten des Experiments zu unvollständig, um auf ihrer Grundlage begründete Aussagen über den Charakter einer der elementaren Verteilungen machen zu können. Wenn aber die Daten, die man als Ergebnis der Modellierung erhalten will, nur wenig von den Fehlern dieser Verteilung abhängen, dann kann die Aufgabe dennoch erfolgreich gelöst werden. Bei der Analyse des Beitrags der Fehler der elementaren Verteilungen zum Gesamtfehler des Modells kann mit Erfolg der Apparat der Varianzanalyse angewendet werden, der oft in der Statistik bei der Aufbereitung experimenteller Daten benutzt wird (Scheffé [1]).

Wie wir schon bemerkt haben, erfordert schließlich die Schätzung von Wahrscheinlichkeiten seltener Ereignisse spezielle Methoden. Die geringe Ordnung der Fehlerabnahme bei der Monte-Carlo-Methode gestattet es nicht, eine kleine Wahrscheinlichkeit mit einem annehmbaren relativen Fehler zu schätzen (bei vernünftigem Arbeitszeitaufwand des Rechenautomaten). In solchen Fällen muß ein gewisses fiktives Modell benutzt werden. Ein einfaches Beispiel dafür wurde im Zusammenhang mit den Aufgaben des Strahlendurchgangs in § 1 besprochen. Nicht für jede reale Aufgabe kann man so leicht ein fiktives Modell mit den nötigen Eigenschaften angeben (und auch begründen). Da bei der Benutzung eines beliebigen Modells die Berechnung der verschiedenen Mittelwerte dem Wesen nach eine Schätzung von Erwartungswerten ist, d. h. von Lebesgueschen Integralen über gewissen σ-endlichen Maßen, können als Quelle fiktiver Modelle gewisse spezielle Integrationsmethoden dienen. Verfahren der Dispersionsverkleinerung, die in den folgenden Kapiteln besprochen werden, haben eine unmittelbare Beziehung zu diesen Fragen.

IV. METHODEN ZUR NÄHERUNGSWEISEN INTEGRALBERECHNUNG

§ 1. Quadraturformeln in Funktionenklassen und die Monte-Carlo-Methode

Wie schon im ersten Kapitel bemerkt wurde, ist die Monte-Carlo-Methode in erster Linie eine Methode zur näherungsweisen Integration. Wenn wir den Erwartungswert irgendeiner Zufallsgröße schätzen, die auf einem Wahrscheinlichkeitsraum $(\mathfrak{X}, \mathfrak{A}, \mu)$ definiert ist, dann schätzen wir das Lebesgue-Stieltjes-Integral

$$I = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})\,. \tag{4.1.1}$$

Soll die Monte-Carlo-Methode im engeren Sinn (Aufgabe A) angewendet werden, so muß dabei $f(x)$ mindestens quadratisch integrierbar sein:

$$\int f^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) < \infty\,.$$

In diesem Kapitel werden hauptsächlich solche Methoden untersucht, die entweder keinerlei Einschränkungen außer der Konstruktivität bezüglich des Maßes μ erfordern oder die sich auf den Fall beziehen, daß für das Lebesguesche Maß eine Wahrscheinlichkeitsdichte $p(\boldsymbol{x})$ existiert. Dabei wird stets die Aufgabe A betrachtet, also die Monte-Carlo-Methode im engeren Sinn, was nicht in jedem konkreten Fall speziell gesagt wird. Die Variable $\boldsymbol{x}$ wird immer als Vektorgröße $\boldsymbol{x} = (x_1, \ldots, x_s)$ aufgefaßt. Theoretisch könnte der Raum $\mathfrak{X}$ als unendlichdimensional angenommen werden (die Endlichkeit von s wird, falls nötig, besonders erwähnt), aber für die Konstruktion numerischer Verfahren im unendlichdimensionalen Fall sind zusätzliche Untersuchungen und Voraussetzungen notwendig (Tschenzow [1], Sobol [3]).

Wenn die Dichte $p(\boldsymbol{x})$ existiert, s endlich ist und $f(\boldsymbol{x})$ einer genügend engen Funktionenklasse angehört (z. B. genügend oft differenzierbar nach allen Variablen ist), kann für die Berechnung des Integrals I der klassische Apparat der Quadraturformeln (bzw. Kubaturformeln) angewendet werden. Für das weitere ist es notwendig, kurz auf einige Fakten und Definitionen aus der Theorie der näherungsweisen Quadratur einzugehen. Dabei verfolgen wir aber keineswegs das Ziel, einen irgendwie vollständigen Überblick über die verschiedenen tiefgehenden Untersuchungen zu diesen Problemen zu geben.

Es sei $p(\boldsymbol{x}) \neq 0$ in einem gewissen Gebiet $\mathcal{D}$ des s-dimensionalen euklidischen Raumes, $\varphi_1(\boldsymbol{x})$ eine in $\mathcal{D}$ definierte Funktion und $f(\boldsymbol{x}) \in \mathbf{F}$ aus einer gewissen Klasse $\mathbf{F}$ von in $\mathcal{D}$ definierten Funktionen, für die das Integral

$$J = \int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, p(\boldsymbol{x})\, f(\boldsymbol{x})\, d\boldsymbol{x}$$

existiert und endlich ist. Gewöhnlich nennt man

$$\int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, p(\boldsymbol{x})\, f(\boldsymbol{x})\, d\boldsymbol{x} \approx \sum_{i=1}^{n} A_i f(\boldsymbol{x}_i) \tag{4.1.2}$$

eine Quadraturformel (im mehrdimensionalen Fall oft Kubaturformel) mit dem Gewicht $p(\boldsymbol{x})$. Die Zahlen A_i, die nicht von der Funktion f abhängen, heißen Koeffizienten (Gewichte) der Quadraturformel, die $\boldsymbol{x}_i$ Knoten (Stützstellen). Häufig wird angenommen, daß alle $\boldsymbol{x}_i \in \mathcal{D}$ sind, was auch im weiteren vorausgesetzt wird.

Wir werden die Summe

$$K_n[f] = \sum_{i=1}^{n} A_i f(\boldsymbol{x}_i) \tag{4.1.3}$$

Näherungssumme nennen und die Größe

$$R_n[f] = \int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x} - \sum_{i=1}^{n} A_i f(\boldsymbol{x}_i) \tag{4.1.4}$$

Restglied der Quadraturformel. Ist für ein $f = f_0$ der Rest $R_n[f_0] = 0$, so sagt man, daß die Quadraturformel (4.1.2) für $f = f_0$ exakt gilt. Es ist leicht zu sehen, daß zu jeder konkreten Funktion $f(\boldsymbol{x})$ eine Quadraturformel mit nur einer Stützstelle konstruiert werden kann, so daß $R_n[f] = 0$ ist. Das setzt allerdings voraus, daß die Größe des Integrals J a priori bekannt ist, und kann daher nicht Gegenstand einer inhaltsreichen Theorie sein. Gewöhnlich wird die Aufgabe der Auswahl der Stützstellen und Koeffizienten (aller oder einiger davon) so gestellt, daß der Rest $R_n[f]$ für alle Funktionen f aus $\mathbf{F}$ genügend klein wird. Eine wichtige Rolle spielt der Begriff der Konvergenz einer Folge von Quadraturformeln. Wir betrachten dazu eine Folge von Näherungssummen

$$K_n[f] = \sum_{i=1}^{n} A_i^{(n)} f(\boldsymbol{x}_i^{(n)}) .$$

Man sagt, daß die Folge der entsprechenden Quadraturformeln für Funktionen der Klasse $\mathbf{F}$ konvergiert, wenn

$$\lim_{n\to\infty} K_n[f] = \int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x}$$

ist. Dabei kann es für jedes n andere Stützstellen $\boldsymbol{x}_i^{(n)}$ und Koeffizienten $A_i^{(n)}$ geben.

Im weiteren werden wir drei Methoden betrachten, eine Quadraturformel zu konstruieren.

1. Es ist ein System von in $\mathcal{D}$ definierten und dort linear unabhängigen Funktionen $\{\varphi_i\}$, $i = 1, \ldots, n$, gegeben (die in (4.1.2) aufgetretene Funktion φ_1 gehört dazu), so daß die Integrale

$$\int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, \varphi_i(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x}\,, \qquad i = 1, \ldots, n\,,$$

bekannt sind. Dann kann man ohne Beschränkung der Allgemeinheit annehmen, daß $\varphi_1(\boldsymbol{x})$ normiert ist, also

$$\int_{\mathcal{D}} \varphi_1^2(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x} = 1$$

ist und die $\varphi_i(\boldsymbol{x})$, $i = 2, \ldots, n$, zu $\varphi_1(\boldsymbol{x})$ orthogonal mit dem Gewicht $p(\boldsymbol{x})$ sind, also

$$\int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, \varphi_i(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x} = 0$$

gilt. Diese Voraussetzung über die $\varphi_i(\boldsymbol{x})$ soll weiterhin immer erfüllt sein.

Wir wählen jetzt n Stützstellen $\boldsymbol{x}_j \in \mathcal{D}$, $j = 1, \ldots, n$, und führen die Bestimmung der Koeffizienten A_i so durch, daß die Quadraturformel exakt für jede Funktion $\varphi_i(\boldsymbol{x})$ gilt. Wurden solche A_i gefunden, so wird die Formel (4.1.2) wegen der Linearität der Aufgabe auch für jede Linearkombination der $\varphi_i(\boldsymbol{x})$ exakt gelten. Schreiben wir die Bedingungen hin, daß (4.1.2) für die $\varphi_i(\boldsymbol{x})$ exakt gelten soll, so ergibt sich zur Bestimmung der A_i ein lineares algebraisches Gleichungssystem

$$\sum_{i=1}^{n} A_i \varphi_j(\boldsymbol{x}_i) = \begin{cases} 1 & \text{für} \quad j = 1\,, \\ 0 & \text{für} \quad j = 2, \ldots, n\,. \end{cases} \tag{4.1.5}$$

Dieses System ist lösbar, wenn für seine Determinante $\Delta(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \det \|\varphi_j(\boldsymbol{x}_i)\| \neq 0$ (i bezeichnet den Zeilenindex) gilt. Das bedeutet eine gewisse Einschränkung bezüglich der Lage der Stützstellen der Quadraturformel. Diese Einschränkungen sind aber sehr geringfügig. Bezeichnet nämlich $\mathcal{D}_n = \mathcal{D} \times \cdots \times \mathcal{D}$ das kartesische Produkt von n Exemplaren des Gebiets $\mathcal{D}$ mit sich selbst, so daß

$$Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathcal{D}_n$$

ist, dann ist leicht zu zeigen, daß das Lebesguesche Maß der Menge $\{Q : \Delta(Q) \neq 0\}$ positiv ist. Das folgt daraus, daß die Gramsche Determinante

$$\det \left\| \int_{\mathcal{D}} \varphi_i(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})\; p(\boldsymbol{x})\, d\boldsymbol{x} \right\|$$

eines Systems orthonormierter Funktionen gleich Eins ist, und aus dem Lemma 4.1, das im nächsten Paragraphen bewiesen wird.

Für die entsprechende Näherungssumme kann man mit der Cramerschen Regel leicht einen expliziten Ausdruck herleiten:

$$K_n[f] = \frac{\Delta(f, Q)}{\Delta Q}. \tag{4.1.6}$$

Dabei ist $\Delta(f, Q) = \det \|f(\boldsymbol{x}_j), \varphi_2(\boldsymbol{x}_j), \ldots, \varphi_n(\boldsymbol{x}_j)\|_1^n$ und $\Delta Q = \Delta(\varphi_1, Q)$.

Gewöhnlich nennt man Formeln der Art (4.1.6) Interpolationsquadraturformeln. Wenn wir erreichen wollen, daß eine Interpolationsquadraturformel für ein Funktionensystem $\{\varphi_i\}$, $i = 1, \ldots, N$ mit $N > n$, exakt gilt, so sind starke Forderungen an die Lage der Stützstellen $\boldsymbol{x}_j$ zu stellen. Jetzt folgt aus (4.1.5) unter der Voraussetzung der Orthogonalität der $\varphi_j(\boldsymbol{x})$, $j = 2, \ldots, N$, zu $\varphi_1(\boldsymbol{x})$, daß die Beziehung

$$\det \|\varphi_r(x_j), \varphi_2(\boldsymbol{x}_j), \ldots, \varphi_n(\boldsymbol{x}_j)\| = 0 \tag{4.1.7}$$

für $n < r \leqq N$ gelten muß. Diese Beziehung stellt ein System nichtlinearer Gleichungen dar, denen die Koordinaten der Punkte $\boldsymbol{x}_j$ genügen müssen. Am vollständigsten ist die Theorie solcher Systeme für den Fall entwickelt, daß die $\varphi_i(\boldsymbol{x})$ Polynome sind (Quadraturformeln vom Gaußschen Typ). Der allgemeine Fall eines beliebigen Systems von Funktionen $\varphi_i(\boldsymbol{x})$, die in einem beliebigen Gebiet $\mathcal{D}$ orthonormiert sind, ist bisher wenig untersucht worden. Jeder konkrete Fall erfordert spezielle Untersuchungen, die besonders im mehrdimensionalen Fall recht kompliziert sind. Im allgemeinen Fall ist es auch schwer, irgendwelche genügend einfache Empfehlungen für die Wahl der Stützstellen einer Interpolationsquadraturformel zu geben, die einen kleinen Fehler für die Funktionen einer Klasse sichern. Das Restglied einer Interpolationsquadraturformel wird häufig durch die Ableitungen des Integranden ausgedrückt (wenn diese existieren). Die entsprechenden Formeln findet man beispielsweise in dem Handbuch von W. I. Krylow [1].

2. Eine zweite Art der Behandlung ist mit der Konstruktion optimaler Quadraturformeln in Klassen differenzierbarer Funktionen verbunden. Am einfachsten läßt sich diese Methode an einem konkreten Beispiel illustrieren. Die Funktion $f(x)$ einer Veränderlichen soll auf dem Intervall $[a, b]$ eine Ableitung haben, die einem gewissen Banachraum $\mathbf{B}$ angehört. Dann kann $f(x)$ auf $[a, b]$ in der Form

$$f(x) = f(a) + \int_a^x df(t)$$

dargestellt werden. Soll die Formel (4.1.2) für eine Konstante exakt sein, so ist ihr Restglied

$$R_n[f] = \int_a^b \int_a^x \frac{df(t)}{b-a}\, dx - \sum_{i=1}^n A_i \int_a^{x_i} df(t)$$

für $\varphi_1(x) = 1$, $p(x) = 1/(b-a)$ ein Element des zu $\mathbf{B}$ adjungierten Raumes der Funktionale, und es läßt sich dessen Norm berechnen. Ist also $f'(t)$ eine stetige Funktion, so gilt

$$|R_n[f]| \leqq \max_{a \leqq x \leqq b} |f'(x)| \int_b^a \left| \frac{b-t}{b-a} - \sum_{i=1}^n A_i\, \vartheta(t - x_i) \right| dt \tag{4.1.8}$$

mit

$$\vartheta(t-x) = \begin{cases} 1 & \text{für} \quad t > x\,, \\ 1/2 & \text{für} \quad t = x\,, \\ 0 & \text{für} \quad t < x\,, \end{cases} \tag{4.1.9}$$

und das Gleichheitszeichen wird für ein bestimmtes f mit stetiger erster Ableitung f' angenommen. Wählt man die A_i und x_i so, daß die Größe

$$\int_a^b \left| \frac{b-t}{b-a} - \sum_{i=1}^n A_i \vartheta(t - x_i) \right| dt$$

minimal wird, dann ergibt sich eine Quadraturformel, die man üblicherweise die beste in der Klasse der einmal stetig differenzierbaren Funktionen nennt. Man

könnte auch sagen, daß sie die beste für die „schlechtesten" Funktionen dieser Klasse ist. Ein analoges Vorgehen ist auch für andere Klassen differenzierbarer Funktionen möglich.

In allen Fällen ergibt sich die Aufgabe, das Minimum einer komplizierten Funktion mehrerer Veränderlicher zu finden. Für gewisse einfachste Funktionenklassen bei einer Veränderlichen kann eine solche Aufgabe vollständig gelöst werden. Besonders schwierig ist das Problem, eine optimale Quadraturformel in Funktionenklassen bei mehreren Veränderlichen zu bestimmen. Wesentliche Beiträge dazu enthalten die Arbeiten von S. L. Sobolew [2]. Für Funktionen mit geringen Differenzierbarkeitseigenschaften erhielt I. M. Sobol [2] einige Resultate, aber eine ausführliche Darstellung dieser Ergebnisse würde den Rahmen dieses Buches sprengen.

3. Die dritte Methode ist mit dem Studium des asymptotischen Verhaltens des Restgliedes $R_n[f]$ der Quadraturformel verknüpft, wenn n gegen Unendlich strebt. Man versucht, für eine gegebene Funktionenklasse solche Formeln zu konstruieren, für die die Abnahme des Fehlers nahezu optimal schnell erfolgt. Dieses Vorgehen hängt mit der vorhergehenden Methode eng zusammen und seine Hervorhebung ist in gewisser Weise willkürlich.

Gut bekannt ist das folgende Resultat (Sobolew [2]). Ist f eine auf dem Intervall $[0, 1]$ (andere Intervalle werden durch eine lineare Transformation auf $[0, 1]$ zurückgeführt) definierte m-mal differenzierbare Funktion einer Veränderlichen, so gibt es eine solche Folge von Quadraturformeln, für deren Restglieder R_n die Abschätzung

$$|R_n[f]| \leqq A_m M_m n^{-m} \tag{4.1.10}$$

gilt. Dabei ist n die Anzahl der Stützstellen, $M_m = \sup\limits_{0 \leqq x \leqq 1} |f^{(m)}(x)|$ und A_m eine Konstante, die nur von den Eigenschaften der Funktion f abhängt.

Ist $f(\boldsymbol{x})$ im s-dimensionalen Einheitswürfel $\mathcal{D}_s$ (oder in einem rechtwinkligen Parallelepiped) definiert, so kann das Integral $\int\limits_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x}$ als iteriertes Integral dargestellt werden, und für seine Berechnung können eindimensionale Quadraturformeln bezüglich jeder Veränderlichen benutzt werden. Besitzt daher die Funktion f in $\mathcal{D}_s$ alle möglichen partiellen Ableitungen m-ter Ordnung und ist n die Anzahl der Stützstellen in der mehrfachen Quadraturformel, so kann eine solche Folge von Formeln gefunden werden, für deren Restglieder R_n die Abschätzung

$$|R_n[f]| \leqq A_{m,s} M_m n^{-m/s} \tag{4.1.11}$$

mit

$$M_m = \sup_{x \in \mathcal{D}} \left| \frac{\partial^m f(\boldsymbol{x})}{\partial x_1^{m_1} \cdots \partial x_s^{m_s}} \right|, \qquad m_1 + \cdots + m_s = m\,,$$

gilt. $A_{m,s}$ ist eine Konstante, die von s und den Eigenschaften der Funktion f abhängt. Dabei kann man solche Funktionen f angeben, für die (4.1.10) bzw. (4.1.11) mit dem Gleichheitszeichen erfüllt sind.

Die Benutzung iterierter Quadraturformeln führt daher mit wachsendem s zu einer starken Verlangsamung der Fehlerabnahme. Im Jahre 1957 zeigte N. S. Bachwalow, daß dieses Resultat mit keiner Integrationsmethode verbessert

werden kann, wenn man nur Informationen über die Zugehörigkeit der Funktion zu einer bestimmten Klasse oder Informationen über Werte der Funktion f und ihrer Ableitungen in einzelnen Punkten benutzt. Genauer wurde von ihm gezeigt: Es sei $\mathbf{H}(m, a, \lambda)$ die Funktionenklasse, für welche die partiellen Ableitungen bis zur m-ten Ordnung einschließlich dem Betrag nach durch die Konstante a beschränkt und die Ableitungen m-ter Ordnung mit dem Exponenten λ $(0 < \lambda \leqq 1)$ hölderstetig sind, also

$$\left| \frac{\partial^m f(\boldsymbol{x}')}{\partial x_1^{m_1} \cdots \partial x_s^{m_s}} - \frac{\partial^m f(\boldsymbol{x}'')}{\partial x_1^{m_1} \cdots \partial x_s^{m_s}} \right| \leqq a \sum_{q=1}^{s} |x_q' - x_q''|^\lambda . \tag{4.1.12}$$

Ist dann $f \in \mathbf{H}(m, a, \lambda)$, so gilt für eine beliebige Integrationsmethode

$$\sup_{f \in \mathbf{H}} |R_n[f]| \geqq A \cdot a_{s, m+\lambda} n^{-(m+\lambda)/s} . \tag{4.1.13}$$

Die Abnahme der Konvergenzordnung mit wachsender Dimensionszahl ist daher durch die umfangreicheren Funktionenklassen erklärbar und nicht durch Unzulänglichkeiten bei der Aufstellung der Quadraturformel bedingt. Es ist natürlich, daß bei der Berechnung von Integralen abzählbarer Vielfachheit (konkrete Aufgaben, bei denen Berechnungen solcher Art erforderlich sind, werden in den nächsten Kapiteln betrachtet) Abschätzungen, die nur Differenzierbarkeitseigenschaften des Integranden ausnutzen, zu schlechten Resultaten führen. Tschenzow [1] zeigt in einer Arbeit, daß es für eine beliebige Quadraturformel der Art (4.1.2) unter den analytischen periodischen Funktionen der Gestalt $\sin^2 (x_i - x_j)$ eine solche gibt, für die der Wert der Näherungssumme den Integralwert mit einem Fehler von mehr als 50% wiedergibt.

Zur Gewinnung von Quadraturformeln mit besseren Eigenschaften ist es daher notwendig, engere Funktionenklassen auszusondern, die durch andere Eigenschaften als Glattheit oder sogar Analytizität zu charakterisieren sind. Ein solcher Weg wurde in den Arbeiten von N. M. Korobow und später von N. S. Bachwalow beschritten.

Eines der Resultate von N. M. Korobow soll hier dargestellt werden.

Es sei $\mathcal{D}_s$ der Einheitswürfel $0 \leqq x_i \leqq 1$, $i = 1, \ldots, s$, und $f(\boldsymbol{x})$ eine in $\mathcal{D}_s$ definierte und bezüglich jeder Variablen x_i periodische Funktion. Wir werden annehmen, daß $f(\boldsymbol{x})$ fast überall in $\mathcal{D}_s$ durch seine Fourierreihe darstellbar ist:

$$f(\boldsymbol{x}) \sim \sum_{m_1, \ldots, m_s = -\infty}^{\infty} c(\boldsymbol{m}) \exp\big(2\pi i(\boldsymbol{m}, \boldsymbol{x})\big) .$$

Dabei ist

$$c(\boldsymbol{m}) = c(m_1, \ldots, m_s) = \int_{\mathcal{D}_s} f(\boldsymbol{x}) \exp\big(2\pi i(\boldsymbol{m}, \boldsymbol{x})\big)\, d\boldsymbol{x}$$

mit

$$(\boldsymbol{m}, \boldsymbol{x}) = \sum_{i=1}^{s} m_i x_i .$$

Weil

$$\int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x} = c(0, \ldots, 0)$$

ist, kann das Restglied der Quadraturformel

$$\int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x} \approx n^{-1} \sum_{k=1}^{n} f(\boldsymbol{x}_k) \tag{4.1.14}$$

in der Form

$$R_n[f] = \sum_{m_1,\ldots,m_s=-\infty}^{\infty}{}' c(\boldsymbol{m})\, n^{-1} \sum_{k=1}^{n} \exp\left(-2\pi i(\boldsymbol{m}, \boldsymbol{x}_k)\right) \tag{4.1.15}$$

geschrieben werden. Dabei bedeutet der Strich beim Summenzeichen das Fehlen des Summanden für $m_1 = \cdots = m_s = 0$. Man kann hoffen, daß das Restglied $R_n[f]$ gemeinsam mit der trigonometrischen Summe

$$S(\boldsymbol{m}) = n^{-1} \sum_{k=1}^{n} \exp\left(2\pi i(\boldsymbol{m}, \boldsymbol{x}_k)\right) \tag{4.1.16}$$

klein wird. Es sei $f(x) \in \mathrm{E}_s^\gamma$ aus der Klasse der in jeder Veränderlichen x_j periodischen Funktionen, für die die Abschätzung

$$c(\boldsymbol{m}) = O\left((\overline{m}_1 \overline{m}_2 \cdots \overline{m}_s)^{-\gamma}\right)$$

erfüllt ist. γ ist eine reelle Zahl größer als Eins, die asymptotische Abschätzung gilt gleichmäßig für die $\overline{m}_i$, und es ist

$$\overline{m}_l = \begin{cases} 1 & \text{für} \quad m_l = 0\,, \\ |m_l| & \text{sonst} \end{cases} \qquad l = 1, \ldots, s\,.$$

Wir wählen einen Punkt $\boldsymbol{x}_k = (x_1^{(k)}, \ldots, x_s^{(k)})$ und setzen

$$x_l^{(k)} = \left\{\frac{a_l k}{n}\right\},$$

wobei $\{a\}$ den gebrochenen Teil der Zahl a bedeutet. Die ganzen, zu n teilerfremden Zahlen $a_1, a_2, \ldots, a_s$ werden wir optimale Koeffizienten nennen, wenn Konstanten $\beta = \beta(s)$ und $c_0 = c_0(s)$ derart existieren, daß für irgendeine Folge von n-Werten die Ungleichung

$$\sum_{m_1,\ldots,m_s=-(n-1)}^{n-1} \frac{\delta_n(a_1 m_1 + a_2 m_2 + \cdots + a_s m_s)}{\overline{m}_1 \overline{m}_2 \cdots \overline{m}_s} \leq c_0 \frac{\log^\beta n}{n} \tag{4.1.17}$$

mit

$$\delta_n(m) = \begin{cases} 1\,, & \text{wenn } n \text{ Teiler von } m\,, \\ 0\,, & \text{wenn } n \text{ kein Teiler von } m\,, \end{cases}$$

erfüllt ist. Die Zahl β heißt Index der optimalen Koeffizienten. Dann gilt der

Satz (Korobow [2]). *Es sei $n > 2$, $a_1, \ldots, a_s$ seien optimale Koeffizienten, $a_i = a_i(n)$, und β der zugehörige Index. Dann gilt für den Fehler der Quadraturformel*

$$\int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_s)\, dx_1 \cdots dx_s = \frac{1}{n} \sum_{k=1}^{n} f\left(\left\{\frac{a_1 k}{n}\right\}, \ldots, \left\{\frac{a_s k}{n}\right\}\right) + R_n[f] \tag{4.1.18}$$

die Abschätzung

$$|R_n[f]| \leqq C n^{-\gamma} \log^{\beta\gamma} n \tag{4.1.19}$$

mit einer nur von γ und s abhängigen Konstanten C.

Wir bemerken, daß man mit der Summe der geometrischen Reihe leicht die Gleichung

$$\delta_n(m) = \frac{1}{n} \sum_{k=1}^{n} \exp\left(2\pi i \frac{mk}{n}\right) \tag{4.1.20}$$

erhält. Aus (4.1.15), (4.1.17) und (4.1.20) folgt dann

$$R_n[f] = \sum_{m_1,\dots,m_s=-\infty}^{\infty} c(\boldsymbol{m})\, \delta_n(\boldsymbol{a}, \boldsymbol{m}) \leqq c \sum_{m_1,\dots,m_s=-\infty}^{\infty} \delta_n(\boldsymbol{a}, \boldsymbol{m})\, (\overline{m}_1 \cdots \overline{m}_s)^{-\beta},$$

wobei $\boldsymbol{a} = (a_1, \dots, a_s)$ ist. Die Optimalität der Koeffizienten $a_1, \dots, a_s$ besteht darin, daß $\delta_n(\boldsymbol{a}, \boldsymbol{m})$ für eine genügend große Anzahl von Werten $m_1, \dots, m_s$ verschwindet. Dann wird der Hauptteil des Restgliedes mit Hilfe der Ungleichung (4.1.17) abgeschätzt, und der Einfluß der übrigen Summanden ist im Verhältnis zum Hauptteil klein.

Eine ausführlichere Darstellung der Fakten, die mit den optimalen Koeffizienten zusammenhängen, kann man in der Monographie von Korobow [2] finden. Dort gibt es insbesondere auch eine Tabelle der optimalen Koeffizienten. Für die Klasse $\mathbf{E}_s^\gamma$ liegt die Abschätzung (4.1.19) nahe bei einer unteren Abschätzung des Integrationsfehlers. Bachwalow bewies, daß für diese Klasse

$$|R_n[f]| \geqq C_1 n^{-\gamma}$$

gilt, wobei C_1 nicht von n abhängt (vgl. auch Scharygin [1]).

4. Die Monte-Carlo-Methode schätzt in ihrer einfachsten Form den Wert des Integrals $\int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$ mit Hilfe des arithmetischen Mittels

$$\frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i), \tag{4.1.21}$$

wobei die $\boldsymbol{x}_i$ unabhängig sind und ihre Verteilung durch das Maß μ bestimmt wird. Formal besitzt diese Formel eine Ähnlichkeit mit den oben beschriebenen Methoden der Integralberechnung. Dem Wesen nach unterscheidet sich die Monte-Carlo-Methode allerdings davon, weil sie auf einem wahrscheinlichkeitstheoretischen Zugang beruht und ihre Konvergenz der Wahrscheinlichkeit nach erfolgt. Infolgedessen gelten folgende Aussagen.

a) Das Restglied

$$\int f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) - \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i)$$

nimmt unabhängig von der Dimension s des Raumes $\mathfrak{X}$ ab. Wie schon im ersten Kapitel bemerkt, wird das Restglied für eine feste Funktion f aus $\mathbf{L}_2(\mu)$ und bei festem Konfidenzniveau γ von der Ordnung $N^{-1/2}$ klein. Übrigens gilt dieselbe Ordnung des Restgliedes auch für eine beliebige Funktion f aus einer Funktionen-

klasse $\mathbf{F}$ mit $\mathbf{F} \subset \mathbf{L}_2(\mu)$. Der Charakter einer Klasse $\mathbf{F}$ mit $\mathbf{F} \subset \mathbf{L}_2(\mu)$ kann sich nur darauf auswirken, wie schnell die Verteilung des arithmetischen Mittels (4.1.21) gegen die Normalverteilung konvergiert. Ist dagegen $\mathbf{L}_2(\mu) \subset \mathbf{F}$, so nimmt das Restglied langsamer als $N^{-1/2}$ ab, und die Grenzverteilung des arithmetischen Mittels ist von der Normalverteilung verschieden.

b) Die Schätzung des Integrals erfolgt sukzessiv. Bei Vergrößerung von N werden in natürlicher Weise die vorher berechneten Funktionswerte mit verwendet. Die deterministischen Quadraturformeln besitzen bei weitem nicht immer diese Eigenschaft.

c) Parallel zur Integralberechnung kann in einfacher Weise der Verfahrensfehler geschätzt werden. Konvergiert also die Verteilung des arithmetischen Mittels mit der Ordnung $O(N^{-1/2})$ gegen die Normalverteilung, dann kann durch Berechnung des arithmetischen Mittels der Quadrate der Funktionswerte in den Punkten $\boldsymbol{x}_i$ die Dispersion von (4.1.21) geschätzt und ein Konfidenzintervall zum gegebenen Konfidenzniveau konstruiert werden. Praktisch erscheint die Fehlerabschätzung für deterministische Quadraturformeln in der Mehrzahl der Fälle als eine hoffnungslos schwierige Aufgabe, wenn f genügend kompliziert ist, denn man muß das Supremum der Ableitungen abschätzen. Schon die Definition einer Klasse differenzierbarer Funktionen, in der f liegt, ist keineswegs einfach. Die Benutzung der Monte-Carlo-Methode fordert freilich auch Kenntnisse über die Zugehörigkeit von f zu einer bestimmten Funktionenklasse. Es ist wünschenswert, daß das Integral $\int |f^3(\boldsymbol{x})| \, \mu(d\boldsymbol{x})$ einen endlichen Wert hat, damit die asymptotische Normalverteilung von (4.1.21) und eine für praktische Zwecke genügend genaue Schätzung der Dispersion gesichert ist. Die Forderung nach der Endlichkeit des dritten Momentes ist allerdings weniger einschränkend als die Beschränktheit der Ableitungen und kann während der Rechnung kontrolliert werden.

Im weiteren werden wir hauptsächlich den Fall $\mathbf{F} \subseteq \mathbf{L}_2(\mu)$ betrachten. Der Fall $\mathbf{F} = \mathbf{L}_p(\mu)$, $1 < p \leqq 2$, wurde bisher wesentlich weniger untersucht. Er wird beispielsweise in den Arbeiten von Kalos [1], Törn [1] und Bachwalow [3] betrachtet (siehe auch Ungleichung (1.3.14) in Kap. I).

Wie schon bemerkt, bleibt bei Benutzung des arithmetischen Mittels (4.1.21) für die Schätzung des Integrals auch im Fall einer engen Funktionenklasse $\mathbf{F}$, $\mathbf{F} \subset \mathbf{L}_2(\mu)$, die Ordnung des Fehlers bei $N^{-1/2}$. Man kann sich aber auf einen anderen Standpunkt stellen und den Fall betrachten, daß es eine gewisse Menge von deterministischen Quadraturformeln gibt, die für die Integration von Funktionen aus $\mathbf{F}$ geeignet sind, und eine konkrete Formel zufällig aus dieser Menge ausgewählt wird. Dieser Weg wurde von N. S. Bachwalow in einigen Arbeiten vorgeschlagen, von denen wir auf die Arbeiten [1], [2] und [3] hinweisen, aber dann untersuchten auch andere Autoren dieses Problem. Die Bildung eines Zufallsmodells solcher Art bietet bestimmte Vorteile, die mit einer Vergrößerung der Konvergenzgeschwindigkeit des Quadraturverfahrens zusammenhängen. Als Beispiel wird die Schätzung des Erwartungswertes des Restgliedes der einfachsten Quadraturformel angeführt:

$$\int_0^1 f(x)\, dx = \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k}{n}\right) + R_n[f] \,. \tag{4.1.22}$$

Bezüglich der Funktion $f(x)$ nehmen wir an, daß ihre Fourierkoeffizienten $c(m)$ $\left(f(x) \sim \sum_{m=-\infty}^{\infty} c(m) \exp(2\pi i m x)\right)$ der Ungleichung

$$\sum_{m=-\infty}^{\infty} |c(m)|^2 m^{2\gamma} \leqq c^2 < \infty$$

genügen. Die so definierte Funktionenklasse bezeichnen wir mit $\mathbf{F}_\gamma(c)$. Die Näherungssumme $K_n[f]$, die der Formel (4.1.22) entspricht, besitzt die leicht nachzuweisende Eigenschaft

$$K_n(e^{2\pi i m x}) = \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i k m/n} = \delta_n(m) = \begin{cases} 0\,, & \text{falls} \quad n \text{ kein Teiler von } m\,, \\ 1\,, & \text{falls} \quad n \text{ Teiler von } m\,. \end{cases}$$

Daraus folgt

$$R_n[f] = \sum_{j=1}^{\infty} [c(nj) + c(-nj)] = \sum_{j=1}^{\infty} c_{nj}$$

mit $c_{nj} = c(nj) + c(-nj)$.

Es ist leicht zu sehen, daß $n^{-\gamma} c \sqrt{2} \cos(2\pi n x) \in \mathbf{F}_\gamma(c)$ ist und folglich

$$\sup_{f \in \mathbf{F}_\gamma(c)} |R_n[f]| \geqq n^{-\gamma} c \sqrt{2} \tag{4.1.23}$$

gilt. Nun seien n und N ganze positive Zahlen mit $N + 1 \leqq n \leqq 2N$ und $\gamma > 1$. Dann kann $R_n[f]$ in der Form

$$R_n[f] = n^{-\gamma} \sum_{j=1}^{\infty} [c_{nj} (nj)^\gamma j^{-\gamma}] = n^{-\gamma} \sum_{j=1}^{\infty} [c_{nj} (nj)^\gamma j^{-1/2}] \cdot [j^{1/2-\gamma}]$$

dargestellt werden. Die Ungleichung von SCHWARZ-BUNJAKOWSKI liefert dann folgende Abschätzung für $|R_n[f]|^2$:

$$|R_n[f]|^2 \leqq n^{-2\gamma} \sum_{j=1}^{\infty} |c_{nj}|^2 (nj)^{2\gamma} j^{-1} \sum_{j=1}^{\infty} j^{1-2\gamma}\,.$$

Die Summe $\sum_{j=1}^{\infty} j^{1-2\gamma}$ ist der Wert der ζ-Funktion im Punkt $1 - 2\gamma$, also gilt

$$|R_n[f]|^2 \leqq n^{-2\gamma} \zeta(1 - 2\gamma) \sum_{j=1}^{\infty} |c_{nj}|^2 (nj)^{2\gamma} j^{-1}\,. \tag{4.1.24}$$

Wir nehmen an, daß n die Werte von $N + 1$ bis $2N$ mit der gleichen Wahrscheinlichkeit $p = 1/N$ annimmt, und schätzen den Erwartungswert des Betrages des Restgliedquadrates ab:

$$\mathsf{M}|R_n^2[f]| = N^{-1} \sum_{n=N+1}^{2N} |R_n[f]|^2\,.$$

Aus (4.1.24) folgt die Ungleichung

$$N^{-1} \sum_{n=N+1}^{2N} n^{2\gamma} |R_n[f]|^2 \leqq \zeta(2\gamma - 1) N^{-1} \sum_{n=N+1}^{2N} \sum_{j=1}^{\infty} \{|c_{nj}|^2 (nj)^{2\gamma} j^{-1}\}\,.$$

Wir setzen $nj = k$. Dann gilt

$$\sum_{j=1}^{\infty} |c_{nj}|^2 (nj)^{2\gamma} j^{-1} = \sum_{k=n}^{\infty}{}'' |c_k|^2 k^{2\gamma} \left(\frac{k}{n}\right)^{-1},$$

wobei die beiden Striche am Summenzeichen bedeuten, daß nur die Summanden vorkommen, für die k ein Vielfaches von n ist. Das ergibt die Abschätzung

$$\sum_{n=N+1}^{2N} \sum_{k=n}^{\infty}{}'' |c_k|^2 k^{2\gamma} \left(\frac{k}{n}\right)^{-1} \leqq \sum_{k=N+1}^{\infty} |c_k|^2 k^{2\gamma} \left(\frac{k}{2N}\right)^{-1} \tau(k; N+1, 2N),$$

wobei $\tau(k; N + 1, 2N)$ die Anzahl der Teiler von k unter den Zahlen von $N + 1$ bis $2N$ bedeutet. Bezeichnet (n, m) den größten gemeinsamen Teiler von n und m, so sind offensichtlich die Aussagen $(n, m) = m$ und $(n, n/m) = n/m$ einander äquivalent. Dann ergibt sich

$$\tau(k; N+1, 2N) = \tau\left(k; \frac{k}{2N}, \frac{k}{N} + 1\right) \leqq \left[\frac{k}{N+1} - \frac{k}{2N}\right] + 1 \leqq \frac{3k}{2N}, \quad k \geqq N.$$

Daraus folgt

$$N^{-1} \sum_{n=N+1}^{2N} n^{2\gamma} |R_n[f]|^2 \leqq \zeta(2\gamma - 1) N^{-1} \sum_{k=N+1}^{\infty} |c_k|^2 k^{2\gamma} \left(\frac{k}{2N}\right)^{-1} \left(\frac{3k}{2N}\right)$$
$$\leqq 6\zeta(2\gamma - 1) c^2 N^{-1} \tag{4.1.25}$$

wegen

$$\sum_{k=N+1}^{\infty} |c_k|^2 k^{2\gamma} \leqq 2 \sum_{k=-\infty}^{\infty} |c(k)|^2 |k|^{2\gamma}.$$

Es ist also

$$\mathsf{M}|R_n[f]|^2 \leqq N^{-2\gamma-1}\, 6c^2 \zeta(2\gamma - 1). \tag{4.1.26}$$

d. h., die Dispersion des Restgliedbetrages hat die Größenordnung $O(N^{-2\gamma-1})$, während für das Supremum des Betrages die Ungleichung (4.1.23) gilt. Man kann auch Resultate herleiten, die sich auf die Verteilung der Größe $|R_n[f]|$ beziehen. Wir wählen dazu ein festes $\varepsilon > 0$. Ist für r Werte der Zahl n die Ungleichung $n^\gamma |R_n[f]| > \varepsilon$ erfüllt, so gilt

$$rN^{-1} \varepsilon^2 < N^{-1} \sum_{n=N+1}^{2N} n^{2\gamma} |R_n[f]|^2.$$

Für $r \geqq \varepsilon^{-2}\, 6\zeta(2\gamma - 1) c^2$ widerspricht das aber (4.1.26). Bei gleichwahrscheinlicher Wahl von n aus $N + 1 \leqq n \leqq 2N$ und für $Nq = r$, $q < 1$, ist mit der Wahrscheinlichkeit $1 - q = 1 - r/N$ die Ungleichung

$$|R_n[f]| \leqq cq^{-1/2} [12\zeta(2\gamma - 1)]^{1/2} n^{-\gamma-1/2}$$

erfüllt. Das ist eines der einfachsten Resultate, die sich auf die zufällige Auswahl einer Quadraturformel aus einer gewissen Menge von Quadraturformeln beziehen. Eine Reihe anderer Ergebnisse kann man in den Arbeiten von Bachwalow [2], [3] und Scharygin [1] finden.

Eine Abschätzung nach unten für den Erwartungswert des Restgliedes einer Quadraturformel in der Funktionenklasse $\mathbf{H}(m, a, \lambda)$ und in einigen anderen Funktionenklassen erhielt BACHWALOW [1]. Insbesondere gilt bei beliebiger Bildung eines Zufallsmodells die Ungleichung

$$\sup_{f \in \mathbf{H}(m,a,\lambda)} \mathsf{M}|R_n[f]| \geqq caN^{-(m+\lambda)/s-1/2} .$$

Durch Bildung eines Zufallsmodells für die Quadraturformeln von N. M. KOROBOW (vgl. (4.1.14)) mit Methoden, die den oben beschriebenen entsprechen, erhält man Schätzungen für den Erwartungswert, die nahe bei den optimal möglichen liegen.

Wie wir sahen, setzt ein solches Vorgehen nicht voraus, daß die entsprechende Näherungssumme eine erwartungstreue Schätzung des Integrals ist. Außerdem wird die Menge von Quadraturformeln, aus denen die zufällige Auswahl erfolgt, als gegeben vorausgesetzt.

In diesem Buch werden in der Regel erwartungstreue Schätzungen des Integrals betrachtet, und die Funktionenklasse wird nicht explizit definiert. Daher haben wir uns auf die Darstellung des einfachsten Beispiels beschränkt, in dem die Anzahl der Stützstellen zufällig gewählt wird. Erwartungstreue Näherungssummen für Funktionen aus $\mathbf{L}_2(\mu)$ werden ausführlich im nächsten Paragraphen betrachtet.

Zum Schluß soll noch auf eine unmittelbare Beziehung zwischen statistischen Testen, die auf eine endliche Zahlenfolge $x_1, x_2, \ldots, x_N$, $0 \leqq x_i \leqq 1$ für alle i, angewendet werden, und dem Fehler der einfachsten Quadraturformel

$$\int_0^1 f(x)\, dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{4.1.27}$$

hingewiesen werden. Die Ergebnisse, die I. M. SOBOL [4], [6] in dieser Richtung erhielt, erlauben es, die Rolle der statistischen Teste bei der Prüfung der Folge der Stützstellen in (4.1.27) auf Gleichverteilung genauer aufzuklären. Ist $f(x)$ auf $[0, 1]$ von beschränkter Variation, so gilt

$$\int_0^1 f(x)\, dx - \frac{1}{N} \sum_{i=1}^{N} f(x_i) = -\int_0^1 \left[\frac{1}{N} \sum_{i=1}^{N} \vartheta(x_i - x) - x\right] df(x) , \tag{4.1.28}$$

wobei die Funktion ϑ durch (4.1.9) definiert wurde. (Vgl. auch (4.1.8), die Gleichungen (4.1.28) bzw. (4.1.8) lassen sich durch partielle Integration herleiten). Aus (4.1.28) folgt

$$|R_n[f] \leqq \sup_{x \in [0,1]} \left| x - N^{-1} \sum_{i=1}^{N} \vartheta(x_i - x) \right| \operatorname{var} f .$$

Die Funktion $G_N(x) = N^{-1} \sum_{i=1}^{N} \vartheta(x_i - x)$ ist die empirische Verteilungsfunktion der Folge $x_1, x_2, \ldots, x_N$. Nimmt man an, daß die zugehörige theoretische Verteilungsfunktion auf $[0, 1]$ eine Gleichverteilung bildet, so stellt der Ausdruck

$$\sqrt{N} \sup_{x \in [0,1]} [x - G_N(x)]$$

die Testgröße von Kolmogoroff für diese Folge dar. Ist also $f(x)$ von beschränkter Variation, dann kann das Restglied bei der Integration unmittelbar durch die Kolmogoroffsche Testgröße ausgedrückt werden. Eine analoge Beziehung kann man auch zum v.-Mises-Smirnow-Test feststellen (Sobol [4]), wenn $f(x)$ eine quadratisch integrierbare Ableitung besitzt. Im mehrdimensionalen Fall kann man analoge Resultate erhalten, wenn man voraussetzt, daß $f(x)$ von beschränkter Variation (im Sinne von Vitali) oder die gemischte Ableitung nach allen Veränderlichen quadratisch integrierbar ist. Eine Darstellung dieser Ideen und Resultate, die die Beziehung zwischen dem Restglied der numerischen Integration und dem Kolmogoroffschen Test betreffen, sind in der Arbeit von Sobol [6] enthalten. Wir werden diese Resultate in Kapitel VII erläutern.

§ 2. Einige allgemeine Methoden zur Dispersionsverkleinerung Stochastische Interpolations-Quadraturformeln.

Durch Benutzung bestimmter Abbildungen der Zufallsgrößen, die den Erwartungswert fest lassen, aber die Dispersion ändern, kann man häufig Modifikationen der Monte-Carlo-Methode gewinnen, die effektiver als die ursprüngliche sind. Die Erhöhung der Effektivität ist in der Regel mit der Auswertung zusätzlicher Informationen über den Integranden und das Maß, bezüglich dessen integriert wird, verbunden. Wir betrachten nun eine Reihe einfacher Modifikationen der Monte-Carlo-Methode, die bei der Integralberechnung Anwendung finden, und untersuchen ihre Besonderheiten.

1. Methode der Abtrennung des Hauptteils. Es soll das Integral $J = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$ berechnet werden, und es gibt eine genügend nahe bei $f(\boldsymbol{x})$ gelegene Funktion $g(\boldsymbol{x})$ derart, daß das Integral $J_1 = \int g(\boldsymbol{x})\, \mu(d\boldsymbol{x})$ leicht analytisch oder mit Hilfe einer einfachen Quadraturformel auszurechnen ist. Dann ist

$$J = J_1 + \int [f(\boldsymbol{x}) - g(\boldsymbol{x})]\, \mu(d\boldsymbol{x}) ,$$

und es genügt, das Integral über die Differenz $\varepsilon(\boldsymbol{x}) = f(\boldsymbol{x}) - g(\boldsymbol{x})$ mit der Monte-Carlo-Methode durch die Summe $\varkappa' = N^{-1} \sum_{i=1}^{N} [f(\boldsymbol{x}_i) - g(\boldsymbol{x}_i)]$ abzuschätzen, wobei die $\boldsymbol{x}_i$ unabhängig und mit der Gesetzmäßigkeit $\mu(d\boldsymbol{x})$ verteilt sind. Für die Dispersion von $\varkappa'$ gilt dann

$$\mathsf{D}\varkappa' = N^{-1} \{\int \varepsilon^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) - [\int \varepsilon(\boldsymbol{x})\, \mu(d\boldsymbol{x})]^2\} . \tag{4.2.1}$$

Bedingungen, unter denen diese Methode vorteilhaft anzuwenden ist, kann man für jede konkrete Aufgabe einfach aufstellen. Zu diesem Zweck genügt es, die Dispersion $\mathsf{D}\varkappa'$ analytisch oder experimentell auf Grund einer Stichprobe mittleren Umfangs zu schätzen.

Wird als $g(\boldsymbol{x})$ eine Linearkombination gewisser gegebener Funktionen gewählt, also

$$g(\boldsymbol{x}) = \sum_{i=1}^{n} a_i \varphi_i(\boldsymbol{x}) ,$$

dann entsteht die Aufgabe, die Konstanten a_i so zu bestimmen, daß die Dispersion $\mathsf{D}\varkappa'$ ein Minimum erreicht. Die Lösung ist in diesem Fall wohl bekannt. Die Funktion $g(\boldsymbol{x})$ muß ein Abschnitt der Fourierreihe von f nach dem Funktionensystem sein, das man aus den φ_i durch Orthogonalisierung erhält. Nimmt man an, daß die φ_i in $\mathbf{L}_2(\mu)$ orthonormiert sind, so müssen die a_i die Fourierkoeffizienten der Funktion f sein, also

$$a_i = \int f(\boldsymbol{x})\, \varphi_i(\boldsymbol{x})\, \mu(d\boldsymbol{x}) .$$

2. Methode der wesentlichen Stichprobe. Das Integral

$$J = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

soll durch die Summe $\varkappa = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{x}_i)$ abgeschätzt werden, wobei die $\boldsymbol{x}_i$ unabhängige Zufallsgrößen mit der Wahrscheinlichkeitsverteilung $\mu(d\boldsymbol{x})$ sind. In diesem Fall beträgt die Dispersion der Zufallsgröße $\varkappa$

$$\mathsf{D}\varkappa = N^{-1} \int [f(\boldsymbol{x}) - J]^2\, \mu(d\boldsymbol{x}) .$$

Nun sei λ ein beliebiges Wahrscheinlichkeitsmaß derart, daß μ absolut stetig in bezug auf λ ist. Dann gilt

$$J = \int f(\boldsymbol{x})\, \frac{d\mu}{d\lambda}\, \lambda(d\boldsymbol{x}) ,$$

und es kann der Schätzwert

$$\varkappa_\lambda = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{y}_i)\, \frac{d\mu}{d\lambda}\, (\boldsymbol{y}_i)$$

benutzt werden, wobei die $\boldsymbol{y}_i$ mit der Gesetzmäßigkeit λ verteilt sind. Die Dispersion beträgt jetzt

$$\mathsf{D}\varkappa_\lambda = N^{-1} \int \left[f(\boldsymbol{x})\, \frac{d\mu}{d\lambda}\, (\boldsymbol{x}) - J \right]^2 \lambda(d\boldsymbol{x}) . \tag{4.2.2}$$

So erhalten wir eine Schar erwartungstreuer Schätzungen für das Integral J, die vom Maß λ abhängen.

Ist $f(\boldsymbol{x})$ eine nichtnegative Funktion, so kann man ein $\lambda = \lambda_0$ so auswählen, daß $\varkappa_{\lambda_0}$ die Dispersion Null besitzt. Setzt man nämlich $\varphi(\boldsymbol{x}) = \frac{d\mu}{d\lambda}\, (\boldsymbol{x})$, so kann man sich überzeugen, daß für

$$\varphi(\boldsymbol{x}) = \begin{cases} J/f(\boldsymbol{x}) & \text{für} \quad f(\boldsymbol{x}) > 0 , \\ 0 & \text{für} \quad f(\boldsymbol{x}) = 0 \end{cases}$$

der Ausdruck (4.2.2) verschwindet. Das entspricht der Aussage

$$\lambda_0(d\boldsymbol{x}) = \frac{1}{J}\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) . \tag{4.2.3}$$

Man muß also die Größe des Integrals vorher kennen, um die Schätzung mit verschwindender Dispersion konstruieren zu können.

Ist $\mu(d\boldsymbol{x}) = p(\boldsymbol{x})\,d\boldsymbol{x}$ (es existiert also die Verteilungsdichte), so gilt $\lambda_0(d\boldsymbol{x}) = J^{-1}f(\boldsymbol{x})\,p(\boldsymbol{x})\,d\boldsymbol{x}$. Ist in diesem Fall $\lambda_0(\boldsymbol{x})$ mit einem Fehler bekannt, also

$$\lambda_1(d\boldsymbol{x}) = \frac{1}{J+\bar{\varepsilon}}\big(f(\boldsymbol{x})\,p(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})\big)\,d\boldsymbol{x}\,, \qquad |\varepsilon(\boldsymbol{x})| < f(\boldsymbol{x})p(\boldsymbol{x})\,,$$

$$\bar{\varepsilon} = \int \varepsilon(\boldsymbol{x})\,d\boldsymbol{x}\,, \qquad \varepsilon(\boldsymbol{x}) = 0 \quad \text{für} \quad f(\boldsymbol{x})\,p(\boldsymbol{x}) = 0\,,$$

dann gilt

$$\mathsf{D}\varkappa_{\lambda_1} = (J+\bar{\varepsilon})\int \frac{f^2(\boldsymbol{x})\,p^2(\boldsymbol{x})}{f(\boldsymbol{x})\,p(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})}\,d\boldsymbol{x} - J^2\,.$$

Nun sei $\varepsilon_0 = \sup\limits_{\boldsymbol{x}} \dfrac{\varepsilon(\boldsymbol{x})}{f(\boldsymbol{x})\,p(\boldsymbol{x})}$. Es ist leicht zu zeigen, daß $\mathsf{D}\varkappa_{\lambda_1} \to J\varepsilon_0^2$ für $\varepsilon_0 \to 0$ gilt. Dieses Resultat, das die Stabilität von $\mathsf{D}\varkappa_\lambda$ in der Umgebung von $\lambda = \lambda_0$ feststellt, kann in gewissen Fällen nützlich sein.

Diese Methode der wesentlichen Stichprobe wird oft in den Fällen benutzt, in denen die Dispersion von $\varkappa$ unendlich ist. Aus (4.2.3) folgt die Möglichkeit, ein solches λ zu wählen, daß (4.2.2) einen endlichen Wert hat, falls nur das Integral $\hat{J} = \int |f(\boldsymbol{x})|\,\mu(d\boldsymbol{x})$ endlich ist. Insbesondere kann man bei der Berechnung des Integrals $\int\limits_{\mathcal{D}} f(\boldsymbol{x})\,d\boldsymbol{x}$ ($\mathcal{D}$ ein Gebiet des euklidischen Raumes), wenn $\int\limits_{\mathcal{D}} f^2(\boldsymbol{x})\,d\boldsymbol{x}$ nicht endlich ist, eine nichtnegative Funktion $\varphi(\boldsymbol{x})$ derart wählen, daß $\int\limits_{\mathcal{D}} \varphi(\boldsymbol{x})\,d\boldsymbol{x} = 1$ ist und $\varkappa_\lambda$ mit $\lambda(d\boldsymbol{x}) = \varphi(\boldsymbol{x})\,d\boldsymbol{x}$ eine endliche Dispersion hat. Man kann beispielsweise

$$\varphi(\boldsymbol{x}) = \begin{cases} \dfrac{|f(\boldsymbol{x})|}{\hat{J}} & \text{für} \quad f(\boldsymbol{x}) \neq 0\,, \\ 0 & \text{für} \quad f(\boldsymbol{x}) = 0 \end{cases}$$

setzen, woraus sich

$$\mathsf{D}\varkappa_\lambda = \Big[\int\limits_{\mathcal{D}} |f(\boldsymbol{x})|\,d\boldsymbol{x}\Big]^2 - \Big[\int\limits_{\mathcal{D}} f(\boldsymbol{x})\,d\boldsymbol{x}\Big]^2 = \hat{J}^2 - J^2$$

ergibt. Ist $\int\limits_{\mathcal{D}} f^2(\boldsymbol{x})\,d\boldsymbol{x}$ nicht endlich, so wird gewöhnlich die Funktion $f(\boldsymbol{x})$ in zwei Faktoren $f = gh$ so zerlegt, daß g keine Singularitäten hat und die Integration von $h > 0$ analytisch durchführbar ist. Sinnvollerweise wird als normierte Funktion h die Verteilungsdichte gewählt, nach der die Integration auszuführen ist. Das Verfahren wird *Abspaltung der Singularität in der Dichte* genannt. Eine Reihe von Anwendungen dazu kann man in dem Buch von BUSLENKO u. a. [1] finden.

Manchmal wird folgende Aufgabe gestellt (EVANS [1]). Es sollen n Integrale $J_i = \int f_i(\boldsymbol{x})\,\mu(d\boldsymbol{x})$ berechnet werden. Wie oben wählt man eine gewisse Verteilungsdichte $\lambda(d\boldsymbol{y})$ und schätzt J_i durch das arithmetische Mittel

$$\varkappa_\lambda^{(i)} = N^{-1}\sum_{j=1}^{N} f_i(\boldsymbol{y}_j)\,\frac{d\mu}{d\lambda}\,(\boldsymbol{y}_j)\,.$$

Zu gegebenen reellen Konstanten a_i, $i = 1, \ldots, n$, wird ein λ derart gewählt, daß die Summe

$$\sum_{i=1}^{n} a_i^2 \, \mathsf{D}\varkappa_\lambda^{(i)} = \int \left(\sum_{i=1}^{n} a_i^2 f_i^2\right) \left(\frac{d\mu}{d\lambda}\right)^2 \lambda(d\boldsymbol{x}) - \sum_{i=1}^{n} a_i^2 J_i^2 \tag{4.2.4}$$

ihr Minimum erreicht. Es genügt, das Minimum von

$$S = \int \left(\sum_{i=1}^{n} a_i^2 f_i^2\right) \left(\frac{d\mu}{d\lambda}\right)^2 \lambda(d\boldsymbol{x})$$

zu suchen. Mit der Bezeichnung

$$\xi = \left(\sum_{i=1}^{n} a_i^2 f_i^2(\boldsymbol{y})\right)^{1/2} \left(\frac{d\mu}{d\lambda}\right)(\boldsymbol{y})$$

ergibt sich $\mathsf{D}\xi = S - \overline{J^2} \geqq 0$ mit

$$\overline{J} = \int \left(\sum_{i=1}^{n} a_i^2 f_i^2\right)^{1/2} \mu(d\boldsymbol{x}), \quad \text{also} \quad S \geqq \left(\int \left(\sum_{i=1}^{n} a_i^2 f_i^2\right)^{1/2} \mu(d\boldsymbol{x})\right)^2. \tag{4.2.5}$$

Das Gleichheitszeichen wird in (4.2.5) in Übereinstimmung mit (4.2.3) für

$$\frac{d\mu}{d\lambda}(\boldsymbol{x}) = \frac{\overline{J}}{\left(\sum_{i=1}^{n} a_i^2 f_i^2(\boldsymbol{x})\right)^{1/2}}$$

angenommen. Also erreicht (4.2.4) sein Minimum bei

$$\lambda(d\boldsymbol{x}) = \overline{J}^{-1} \left(\sum_{i=1}^{n} a_i^2 f_i^2(\boldsymbol{x})\right)^{1/2} \mu(d\boldsymbol{x})$$

und ist gleich

$$\overline{J^2} - \sum_{i=1}^{n} a_i^2 J_i^2 \,.$$

Der letzte Ausdruck verschwindet, wenn die Summe $\sum_{i=1}^{n} a_i^2 f_i^2(\boldsymbol{x})$ konstant ist. Insbesondere verschwindet der Ausdruck (4.2.4), wenn alle $f_i(\boldsymbol{x}) > 0$ und konstant sind. Diese Überlegungen erlauben es, a-priori-Informationen über die Integranden zu berücksichtigen und die Effektivität der Monte-Carlo-Methode zu erhöhen.

Auf die beiden beschriebenen Methoden werden wir im weiteren in Verbindung mit konkreten Aufgaben zurückkommen. Offensichtlich kann man sie kombinieren, indem man etwa zuerst den Hauptteil des Integranden abtrennt und danach die Methode der wesentlichen Stichprobe anwendet.

Als Beispiel für die Anwendung dieser Methoden führen wir das in der Arbeit von Rosenberg [1] beschriebene Integrationsverfahren an, das auf der Anwendung der zum Integranden gehörigen Bernsteinpolynome beruht.

Es sei

$$J = \int_0^1 dx_1 \cdots \int_0^1 dx_s\, f(x_1, \ldots, x_s) = \int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x}\,,$$

wobei $f(\boldsymbol{x})$ eine in $\mathcal{D}_s$ stetige Funktion ist. Bernsteinpolynom der Funktion f wird das Polynom

$$B^f_{m_1,\ldots,m_s}(\boldsymbol{x}) \equiv B^f_{\boldsymbol{m}}(\boldsymbol{x}) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_s=0}^{m_s} f\left(\frac{l_1}{m_1}, \ldots, \frac{l_s}{m_s}\right) \prod_{k=1}^{s} P_{l_k, m_k}(x_k) \tag{4.2.6}$$

mit $\boldsymbol{m} = (m_1, \ldots, m_s)$ und $P_{l,m}(x) = \binom{m}{l} x^l (1-x)^{m-l}$ genannt. Für $m_1, \ldots, m_s \to \infty$ strebt $B^f_{\boldsymbol{m}}(\boldsymbol{x}_0)$ gegen $f(\boldsymbol{x}_0)$ in jedem Stetigkeitspunkt der Funktion f. Besitzt f im Punkt $\boldsymbol{x}_0$ eine stetige partielle Ableitung der Ordnung k, so strebt die entsprechende partielle Ableitung des Bernsteinpolynoms im Punkt $\boldsymbol{x}_0$ ebenfalls gegen den Wert dieser Ableitung (Natanson [1]). Auf diese Weise berücksichtigt $B^f_{\boldsymbol{m}}(\boldsymbol{x})$ die Besonderheiten im Verlauf der Funktion $f(\boldsymbol{x})$ genügend gut.

Wir nehmen nun an, daß $f(\boldsymbol{x}) \geqq 0$ im Einheitswürfel $\mathcal{D}_s$ gilt, über den die Integration durchgeführt wird. Weil das Hinzufügen einer Konstanten zu $f(\boldsymbol{x})$ die Integrationsaufgabe nicht komplizierter macht, kann man annehmen, daß $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) \geqq A > 0$ gilt. Dann ergibt sich

$$B^f_{\boldsymbol{m}}(\boldsymbol{x}) \geqq A \prod_{k=1}^{s} \sum_{l_k=0}^{m_k} P_{l_k, m_k}(x_k) = A > 0\,,$$

weil jede Summe eine Summe über die Wahrscheinlichkeiten der Binomialverteilung darstellt und folglich gleich Eins ist. Wir bemerken auch, daß

$$\int_{\mathcal{D}_s} B^f_{\boldsymbol{m}}(\boldsymbol{x})\, d\boldsymbol{x} \equiv \overline{S} = \frac{S}{\prod\limits_{k=1}^{s} (m_k + 1)}$$

mit

$$S = \sum_{l_1=0}^{m_1} \cdots \sum_{l_s=0}^{m_s} f\left(\frac{l_1}{m_1}, \ldots, \frac{l_s}{m_s}\right)$$

gilt. Wir wenden uns nun der Methode der wesentlichen Auswahl zu und setzen

$$\lambda(d\boldsymbol{x}) = \tilde{B}^f_{\boldsymbol{m}}(\boldsymbol{x})\, d\boldsymbol{x}$$

mit

$$\tilde{B}^f_{\boldsymbol{m}}(\boldsymbol{x}) = \frac{B^f_{\boldsymbol{m}}(\boldsymbol{x})}{\overline{S}}$$

$$= \frac{1}{\overline{S}} \sum_{l_1=0}^{m_1} \cdots \sum_{l_s=0}^{m_s} f\left(\frac{l_1}{m_1}, \ldots, \frac{l_s}{m_s}\right) \prod_{k=1}^{s} \frac{(m_k+1)!}{l_k!\,(m_k-l_k)!} x_k^{l_k} (1-x_k)^{m_k - l_k}\,,$$

so daß sich

$$\int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x} = \overline{S} \int_{\mathcal{D}_s} \frac{f(\boldsymbol{x})}{B^f_{\boldsymbol{m}}(\boldsymbol{x})}\, \tilde{B}^f_{\boldsymbol{m}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.2.7}$$

ergibt.

Eine angenehme Eigenschaft einer solchen Wahl von λ besteht darin, daß sich die entsprechende Zufallsgröße verhältnismäßig einfach modellieren läßt. Die Dichte $\tilde{B}_m^f(\boldsymbol{x})$ ist mit folgendem stochastischem Modell verknüpft. Es seien $M = \prod_{k=1}^{s} (m_k + 1)$ Urnen gegeben, die durch die Indizes $l_1, \ldots, l_s$ gekennzeichnet sind ($l_k = 0, \ldots, m_k$; $k = 1, \ldots, s$). Die Wahrscheinlichkeit, aus einer Urne mit den Indizes $l_1, \ldots, l_s$ zu ziehen, sei gleich $f(l_1/m_1, \ldots, l_s/m_s)/S$. Nach der Wahl einer Urne, d. h. von Werten der Indizes $l_1, \ldots, l_s$, erzeugt man entsprechend der gewünschten Wahrscheinlichkeitsverteilung s unabhängige Zufallsgrößen y_k mit einer β-Verteilung, deren Parameter l_k und $m_k - l_k$ sind. Für $l_k = m_k$ oder $l_k = 0$ erhält man ausgeartete β-Verteilungen, die den Dichten $(m_k + 1)\, x_k^{m_k}$ bzw. $(m_k + 1) \times (1 - x_k)^{m_k}$ entsprechen (vgl. (2.3.17)). Die Komponenten des Vektors $(y_1, \ldots, y_s)$ besitzen die gemeinsame Verteilungsdichte $\tilde{B}_m^f(\boldsymbol{x})$.

Ein Bernsteinpolynom kann natürlich in der Methode der Abtrennung des Hauptteils auch als Funktion $g(\boldsymbol{x})$ gewählt werden. Man kann auch bei der Schätzung des Erwartungswertes der Differenz $f(\boldsymbol{x}) - B_m^f(\boldsymbol{x})$ eine wesentliche Stichprobe mit der Dichte $\tilde{B}_m^f(\boldsymbol{x})$ benutzen, was allerdings in der Regel keinen wesentlichen Vorteil bietet (die Differenz $f(\boldsymbol{x}) - B_m^f(\boldsymbol{x})$ kann im Verhalten andere Eigenschaften besitzen als $f(\boldsymbol{x})$).

Das beschriebene Integrationsverfahren ist dem Wesen nach ein Verfahren, das mit Hilfe der Monte-Carlo-Methode die einfachste Quadraturformel $J \approx \overline{S}$, die man als Ergebnis der wiederholten Anwendung der Rechteckformel erhält, genauer macht. Bei der Berechnung von mehrfachen Integralen hoher Dimension ist seine Anwendung mit Schwierigkeiten verbunden, die im allgemeinen bei der Benutzung äquidistanter Netze auftreten. Nichtsdestoweniger ist dieses Verfahren von prinzipiellem Interesse, da es die klassischen Integrationsmethoden mit der Monte-Carlo-Methode verbinde.

3. Erniedrigung der Integrationsordnung (Buslenko u. a. [1]). Die Veränderlichen $\boldsymbol{x} = (x_1, \ldots, x_s)$ werden in zwei Gruppen eingeteilt: $\boldsymbol{y} = (x_1, \ldots, x_t)$ und $\boldsymbol{z} = (x_{t+1}, \ldots, x_s)$, $t < s$, so daß sich

$$J = \int_{(\boldsymbol{y})} \int_{(\boldsymbol{z})} f(\boldsymbol{y}, \boldsymbol{z})\, \mu(d\boldsymbol{y}\, d\boldsymbol{z})$$

ergibt. Wir nehmen an, daß man

$$\nu(d\boldsymbol{y}) = \int_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z}) \quad \text{und} \quad \tilde{f}(\boldsymbol{y}) = \int_{(\boldsymbol{z})} f(\boldsymbol{y}, \boldsymbol{z})\, \tilde{\mu}(d\boldsymbol{z})$$

analytisch berechnen kann, wobei $\tilde{\mu}(d\boldsymbol{z})$ durch die Beziehung

$$\tilde{\mu}(d\boldsymbol{z}) \int_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z}) = \mu(d\boldsymbol{y}\, d\boldsymbol{z})$$

definiert ist. Dann ergibt sich

$$J = \int_{(\boldsymbol{y})} \tilde{f}(\boldsymbol{y})\, \nu(d\boldsymbol{y})\,.$$

Bei der Abschätzung des letzten Integrals durch die Summe

$$\varkappa'' = N^{-1} \sum_{i=1}^{N} \tilde{f}(\boldsymbol{y}_i) ,$$

wobei die $\boldsymbol{y}_i$ mit der Gesetzmäßigkeit ν verteilt sind, kann die Dispersion nur kleiner werden als bei der Abschätzung von J durch die Summe

$$\varkappa = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{y}_i, \boldsymbol{z}_i) .$$

In der Tat ist

$$\mathsf{D}\varkappa - \mathsf{D}\varkappa'' = \int\limits_{(\boldsymbol{y})} \left\{ \frac{\int\limits_{(\boldsymbol{z})} f^2(\boldsymbol{y},\boldsymbol{z})\, \mu(d\boldsymbol{y}\, d\boldsymbol{z})}{\int\limits_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z})} - \left(\frac{\int\limits_{(\boldsymbol{z})} f(\boldsymbol{y},\boldsymbol{z})\, \mu(d\boldsymbol{y}\, d\boldsymbol{z})}{\int\limits_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z})} \right)^2 \right\} \int\limits_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z}) \geqq 0 ,$$

weil der Ausdruck in den geschweiften Klammern die Dispersion der Zufallsgröße $f(\boldsymbol{y}, \boldsymbol{z})$ bei festem $\boldsymbol{y}$ darstellt. (Die Verteilung von $\boldsymbol{z}$ ist gleich $\mu(d\boldsymbol{y}\, d\boldsymbol{z}) / \int\limits_{(\boldsymbol{z})} \mu(d\boldsymbol{y}\, d\boldsymbol{z})$.)

4. Geschichtete Stichprobe. Wir stellen den Raum $\mathfrak{X}$, über den integriert werden soll, als endliche Summe von m durchschnittsfremden Mengen A_i dar und erhalten

$$\int f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = \sum_{i=1}^{m} \int\limits_{A_i} f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) . \tag{4.2.8}$$

Jedes Integral auf der rechten Seite von (4.2.8) kann mit der gewöhnlichen Monte-Carlo-Methode oder einer ihrer bereits betrachteten Modifikationen geschätzt werden. Wir setzen voraus, daß das Integral

$$\int\limits_{A_i} f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = \mu(A_i) \int g_i(\boldsymbol{x})\, \mu(d\boldsymbol{x}) / \mu(A_i)$$

durch die Summe

$$n_i^{-1}\, \mu(A_i) \sum_{k_i=1}^{n_i} g_i(\boldsymbol{x}_{k_i}) \quad \text{mit} \quad g_i(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{für} \quad \boldsymbol{x} \in A_i , \\ 0 & \text{sonst} . \end{cases}$$

abgeschätzt wird. Dabei sind die $\boldsymbol{x}_{k_i}$, $k_i = 1, \ldots, n_i$; $i = 1, \ldots, m$, Realisierungen der unabhängigen Zufallsgrößen $\boldsymbol{x}_i$ mit einer Wahrscheinlichkeitsverteilung $\mu(d\boldsymbol{x}) / \mu(A_i)$, $\boldsymbol{x} \in A_i$. So ergibt sich

$$\int f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \approx \sum_{i=1}^{m} n_i^{-1}\, \mu(A_i) \sum_{k_i=1}^{n_i} g_i(\boldsymbol{x}_{k_i}) \equiv \hat{\varkappa}[f] .$$

Offensichtlich ist $\mathsf{M}\hat{\varkappa}[f] = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$. Die Dispersion der Summe $\hat{\varkappa}$ läßt sich leicht berechnen. Sie ist gleich

$$\mathsf{D}\hat{\varkappa} = \sum_{i=1}^{m} n_i^{-1}\, \mu(A_i) \int\limits_{A_i} f^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) - \sum_{i=1}^{m} n_i^{-1} \left[\int\limits_{A_i} f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \right]^2 . \tag{4.2.9}$$

Es sei nun $m = 2$, $p_1 = \mu(A_1)$, $p_2 = \mu(A_2)$, $n_1 + n_2 = N$. Wir werden p_1, p_2, n_1, n_2 derart annehmen, daß $n_1 = p_1 N$ und $n_2 = p_2 N$ ganze Zahlen sind. Dann ergibt sich

$$\mathsf{D}\hat{\varkappa} = \frac{1}{N}\left\{\int f^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) - \frac{1}{p_1}\left[\int_{A_1} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]^2 - \frac{1}{p^2}\left[\int_{A_2} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]^2\right\}. \tag{4.2.10}$$

Bei der gewöhnlichen Monte-Carlo-Methode, die für die Schätzung des Integrals über den ganzen Raum $\mathfrak{X}$ angewendet wird, ergibt sich

$$\mathsf{D}\varkappa = \frac{1}{N}\left\{\int f^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) - \left[\int_{A_1} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]^2 - \left[\int_{A_2} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]^2 - 2\left[\int_{A_1} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]\cdot\left[\int_{A_2} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]\right\}.$$

Da $1/p_1 > 1$ und $1/p_2 > 1$ ist, ist leicht zu sehen, daß $\mathsf{D}\hat{\varkappa} \leqq \mathsf{D}\varkappa$ für beliebiges f aus $\mathbf{L}_2(\mu)$ gilt. Gleichheit tritt hier nur ein, wenn

$$\mu(A_2)\int_{A_1} f(\boldsymbol{x})\,\mu(d\boldsymbol{x}) = \mu(A_1)\int_{A_2} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})$$

ist.

So führt eine beliebige Zerlegung, bei der die Anzahl der Punkte proportional zum Maß der Mengen A_1 und A_2 gewählt wird, zu einer Verkleinerung der Dispersion im Vergleich mit der gewöhnlichen Monte-Carlo-Methode. Mit anderen Worten, es gibt keine solche Funktion aus $\mathbf{L}_2(\mu)$, für die die Schichtung mit einer proportionalen Anzahl von Punkten ein schlechteres Resultat liefert, als die gewöhnliche Monte-Carlo-Methode. Wird die Anzahl der Punkte nicht proportional zu den Maßen von A_1 bzw. A_2 gewählt, so gilt diese Aussage im allgemeinen nicht.

Setzt man etwa $m = 2$, $n_1 = n_2$, $\mu(A_1) = 3/4$ und $\mu(A_2) = 1/4$, so erhält man

$$\mathsf{D}\hat{\varkappa} = \frac{1}{n_1}\left\{\frac{3}{4}\int_{A_1} f^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) + \frac{1}{4}\int_{A_2} f^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) - \left(\int_{A_1} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right)^2 - \left(\int_{A_2} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right)^2\right\}.$$

Durch Vergleich dieses Ausdrucks mit der Dispersion der gewöhnlichen Monte-Carlo-Methode bei $2n_1$ Versuchen überzeugt man sich, daß es eine Funktion gibt, für die eine solche geschichtete Auswahl ein schlechteres Resultat liefert als die Schätzung mit der gewöhnlichen Monte-Carlo-Methode (z. B. eine Funktion, die nur auf A_1 ungleich Null ist). Wenn man allerdings über einige zusätzliche Informationen über die Funktion f verfügt, dann kann man n_1 und n_2 so wählen, daß $\mathsf{D}\hat{\varkappa}$ minimal wird (Gelfand, Frolow und Tschenzow [1]).

Aus (4.2.10) erhält man leicht

$$\mathsf{D}\hat{\varkappa} = \frac{1}{p_1 N}\mu^2(A_1)\,\mathsf{D}g_1 + \frac{1}{p_2 N}\,\mu^2(A_2)\,\mathsf{D}g_2\,.$$

Der Wert p_1, für den $\mathsf{D}\hat{\varkappa}$ minimal wird, ist gleich

$$p_1 = \frac{\mu(A_1)\sqrt{\mathsf{D}g_1}}{\mu(A_1)\sqrt{\mathsf{D}g_1} + \mu(A_2)\sqrt{\mathsf{D}g_2}},$$

und die minimale Dispersion ist

$$\mathsf{D}_{\min} = \frac{1}{N}\left[\mu(A_1)\sqrt{\mathsf{D}g_1} + \mu(A_2)\sqrt{\mathsf{D}g_2}\right].$$

5. Stochastische Quadraturformeln. Ein universelles Verfahren zur Verkleinerung der Dispersion bei der Monte-Carlo-Methode ist die Benutzung stochastischer Quadraturformeln. In diesem Paragraphen werden wir Interpolationsquadraturformeln betrachten, deren Stützstellen eine Verteilungsfunktion spezieller Art besitzen. Es ergeben sich auch eine Reihe von Hilfssätzen, die später gebraucht werden. Stochastische Quadraturformel nennen wir eine Formel

$$\sum_{i=1}^{n} A_i f(\boldsymbol{x}_i) \approx \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x}), \tag{4.2.11}$$

wobei die A_i und $\boldsymbol{x}_i$ Zufallsgrößen und φ_1 und f gegebene Funktionen sind, so daß das Produkt $\varphi_1 f$ μ-integrierbar ist. Die Summe $\varkappa_n[f] = \sum_{i=1}^{n} A_i f(\boldsymbol{x}_i)$, die der Formel (4.2.11) entspricht, werden wir stochastische Näherungssumme nennen. Weiterhin werden wir voraussetzen, ohne das besonders hervorzuheben, daß die Summe $\varkappa_n[f]$ eine erwartungstreue Schätzung des Integrals

$$J = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

darstellt und die Koeffizienten gegebene Funktionen der Zufallsgrößen $\boldsymbol{x}_i$ sind:

$$A_i = A_i(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

Als einfachstes Beispiel einer stochastischen Näherungssumme kann die Summe

$$\varkappa_n[f] = \sum_{i=1}^{n} \frac{\varphi_1(\boldsymbol{x}_i)}{n} f(\boldsymbol{x}_i)$$

dienen, die eine erwartungstreue Schätzung des Integrals J darstellt, wenn die $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ unabhängig sind und die Verteilung $\mu(d\boldsymbol{x})$ besitzen.

Gilt für $f = f_0$ mit der Wahrscheinlichkeit Eins die Gleichung

$$\sum_{i=1}^{n} A_i f_0(\boldsymbol{x}_i) = \int \varphi_1(\boldsymbol{x})\, f_0(\boldsymbol{x})\, \mu(d\boldsymbol{x}),$$

so werden wir sagen, daß die Quadraturformel für die Funktion f_0 exakt gilt.

Für das Weitere benötigen wir ein Lemma, das die Verallgemeinerung einer bekannten Integralidentität auf den Fall eines beliebigen σ-endlichen Maßes[1]) darstellt (Andrejew [1]).

[1]) Für die weiteren Betrachtungen ist es bequem anzunehmen, daß das Maß μ endlich oder σ-endlich ist. Das wird in der Regel nicht besonders hervorgehoben. Es muß aber klar sein, daß der Gebrauch des Ausdrucks $\mu(\mathfrak{X})$ ein endliches Maß voraussetzt.

Lemma 4.1. *Es seien* $\{\varphi_i\}$, $i = 1, \ldots, n$, *und* $\{\psi_j\}$, $j = 1, \ldots, n$, *zwei Funktionensysteme aus* $\mathbf{L}_2(\mu)$. *Dann gilt die Integralidentität*

$$I \equiv \int \mu_n(dQ) \det \|\varphi_i(\boldsymbol{x}_j)\| \det \|\psi_k(\boldsymbol{x}_l)\| = n! \det \|(\varphi_i, \psi_k)\| \tag{4.2.12}$$

mit

$$Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) , \qquad \mu_n\,(dQ) = \mu(d\boldsymbol{x}_1) \cdots \mu(d\boldsymbol{x}_n)$$

und

$$(\varphi_i, \psi_k) = \int \varphi_i(\boldsymbol{x})\, \psi_k(\boldsymbol{x})\, \mu(d\boldsymbol{x}) .$$

Beweis. Wir entwickeln die Determinante $\det \varphi_i(x_j)$ (j Nummer der Zeile) nach den Elementen der ersten Spalte, wobei Δ_j den entsprechenden Minor bezeichnet:

$$\Delta_j = \begin{vmatrix} \varphi_2(\boldsymbol{x}_1) & \varphi_3(\boldsymbol{x}_1) & \ldots & \varphi_n(\boldsymbol{x}_1) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_2(\boldsymbol{x}_{j-1}) & \varphi_3(\boldsymbol{x}_{j-1}) & \ldots & \varphi_n(\boldsymbol{x}_{j-1}) \\ \varphi_2(\boldsymbol{x}_{j+1}) & \varphi_3(\boldsymbol{x}_{j+1}) & \ldots & \varphi_n(\boldsymbol{x}_{j+1}) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_2(\boldsymbol{x}_n) & \varphi_3(\boldsymbol{x}_n) & \ldots & \varphi_n(\boldsymbol{x}_n) \end{vmatrix} .$$

Wir erhalten

$$\begin{aligned} I &= \int \mu_n(dQ) \sum_{j=1}^{n} (-1)^{j+1} \Delta_j \varphi_1(\boldsymbol{x}_j) \det \|\psi_k(\boldsymbol{x}_l)\| \\ &= \sum_{j=1}^{n} \int \mu(d\boldsymbol{x}_1) \cdots \int \mu(d\boldsymbol{x}_{j-1}) \int \mu(d\boldsymbol{x}_{j+1}) \\ &\quad \cdots \int \mu(d\boldsymbol{x}_n)\, (-1)^j \Delta_j \int \mu(d\boldsymbol{x}_j)\, \varphi(\boldsymbol{x}_j) \det \|\psi_k(\boldsymbol{x}_l)\| . \end{aligned}$$

Wir multiplizieren die j-te Zeile der Determinante $\det \|\psi_k(\boldsymbol{x}_l)\|$ mit $\varphi_1(\boldsymbol{x}_j)$ und integrieren über ihre Elemente. Die Vertauschung der zu integrierenden Zeile mit der ersten Zeile ändert den Wert der Determinante um den Faktor $(-1)^{j-1}$. Wird jetzt die Numerierung der ersten $j-1$ Integrationsveränderlichen so geändert, daß $\boldsymbol{x}_2$ anstelle von $\boldsymbol{x}_1$, $\boldsymbol{x}_3$ anstelle von $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_j$ anstelle von $\boldsymbol{x}_{j-1}$ gesetzt wird, dann ergibt sich

$$I = n \int \mu(d\boldsymbol{x}_2) \cdots \int \mu(d\boldsymbol{x}_n) \det \|\varphi_i(\boldsymbol{x}_j)\| \begin{vmatrix} (\varphi_1, \psi_1) & (\varphi_1, \psi_2) & \ldots & (\varphi_1, \psi_n) \\ \psi_1(\boldsymbol{x}_2) & \psi_2(\boldsymbol{x}_2) & \ldots & \psi_n(\boldsymbol{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ \psi_1(\boldsymbol{x}_n) & \psi_2(\boldsymbol{x}_n) & \ldots & \psi_n(\boldsymbol{x}_n) \end{vmatrix} .$$

So haben wir die Ordnung der einen Determinante um 1 erniedrigt und eine Zeile der zweiten Determinante integriert. Eine sinngemäße Wiederholung dieses Schrittes führt uns offensichtlich zum gewünschten Resultat. Damit ist das Lemma bewiesen.

Unmittelbar aus dem Lemma ergibt sich folgendes Resultat, das Ermakow und Solotuchin [1] erhielten und das durch Handscomb [1] präzisiert wurde.

Satz 4.1. *Wenn*

1. *die* $\{\varphi_i\}$, $i = 1, \ldots, n$, *für fast alle* $\boldsymbol{x}$ *linear unabhängige Funktionen sind, wobei* φ_1 *orthogonal zu* φ_i, $i = 2, \ldots, n$, *und normiert ist,*

2. *die Verteilung* $W_0(dQ)$ *der Stützstellen der Näherungssumme*

$$\varkappa_n[f] = \frac{\Delta(f; Q)}{\Delta(Q)} \tag{4.2.13}$$

(vgl. (4.1.6)) *durch die Gleichung* $W_0(dQ) = c\,\Delta^2(Q)\,\mu_n(dQ)$ *mit einer Normierungskonstanten c bestimmt wird (für orthonormierte* φ_i *ist* $c = 1/n!$),
dann ist $\varkappa_n[f]$ *eine erwartungstreue Schätzung des Integrals*

$$\int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \equiv (\varphi_1, f)\,.$$

Beweis. Wir setzen zunächst voraus, daß die φ_i, $i = 1, \ldots, n$, orthonormiert sind. In diesem Fall ergibt sich mit $\psi_i \equiv \varphi_i$, $i = 1, \ldots, n$, und durch Benutzung von Lemma 4.1

$$\int \Delta^2(Q)\, \mu_n(dQ) = n!\, \det \|(\varphi_i, \varphi_j)\| = n!\,,$$

also $c = 1/n!$. Nun wird Lemma 4.1 zur Berechnung von $\mathsf{M}\varkappa_n[f]$ benutzt, indem man $\psi_1(\boldsymbol{x}) = f(\boldsymbol{x})$ und $\psi_i = \varphi_i$, $i = 2, \ldots, n$, setzt. Dann erhält man

$$\begin{aligned} \mathsf{M}\varkappa_n[f] &= \frac{1}{n!} \int \frac{1}{\Delta(Q)}\, \Delta(f; Q)\, \Delta^2(Q)\, \mu_n(dQ) \\ &= \frac{1}{n!} \int \Delta(f; Q)\, \Delta(Q)\, \mu_n(dQ) = (\varphi_1, f)\,, \end{aligned}$$

wodurch der Satz für orthonormierte Funktionen φ_i bewiesen ist.

Der Fall, daß $\varphi_1(\boldsymbol{x})$ orthogonal zu den übrigen $\varphi_i(\boldsymbol{x})$ ist und alle φ_i für fast alle $\boldsymbol{x}$ (also auf dem Träger des Maßes μ) linear unabhängig sind, wird leicht auf den vorhergehenden Fall zurückgeführt. Dazu genügt die Feststellung, daß die Multiplikation einer Spalte mit einer Konstanten und die Addition beliebiger Spalten in den Determinanten, die im Zähler und Nenner von (4.2.13) stehen, den Wert von $\varkappa_n[f]$ nicht ändert. Auf diese Weise können alle φ_i orthonormiert werden, wobei die Größen $\Delta(f; Q)$ und $\Delta(Q)$ mit derselben Konstanten multipliziert werden und $\varkappa_n[f]$ sich nicht ändert. So ergibt sich der bereits betrachtete Fall orthonormierter φ_i. Damit ist der Satz vollständig bewiesen.

Weiterhin werden wir auf $\mathfrak{X}$ zwischen regulären und nichtregulären Funktionensystemen $\{\varphi_i\}$, $i = 1, \ldots, n$, unterscheiden. Wir nennen $\{\varphi_i\}$ ein auf $\mathfrak{X}$ bezüglich des Maßes μ *reguläres System*, wenn das Maß μ_n der Menge $\{Q : \Delta(Q) = 0\}$ gleich Null ist. Andernfalls heißt $\{\varphi_i\}$ *nichtregulär* auf $\mathfrak{X}$. Offensichtlich sind die Funktionen eines auf $\mathfrak{X}$ regulären Systems auf einer beliebigen Teilmenge von $\mathfrak{X}$ positiven Maßes fast überall linear unabhängig. Für nichtreguläre Systeme gibt es eine solche Menge $d \in \mathfrak{X}$ positiven Maßes, auf der die φ_i linear abhängig sind. Es gilt der

Satz 4.2 (Ermakow und Solotuchin [1], Ermakow [4]). *Wenn die* φ_i, $i = 1, \ldots, n$, *ein System orthonormierter Funktionen bilden, dann gilt unter den*

Voraussetzungen von Satz 4.1 die Ungleichung

$$\mathsf{D}\varkappa_n[f] \leqq \int \left[f(\boldsymbol{x}) - \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x}) \right]^2 \mu(d\boldsymbol{x}) \,. \tag{4.2.14}$$

Bilden die φ_i ein reguläres System, so gilt in (4.2.14) *das Gleichheitszeichen.*

Beweis. Es ist

$$\mathsf{D}\varkappa_n[f] = \int (\varkappa_n[f])^2 \frac{\Delta^2(Q)}{n!}\, \mu_n(dQ) - (f, \varphi_1)^2 \,.$$

Im Fall eines regulären Systems ergibt sich

$$\begin{aligned} \mathsf{D}\varkappa_n[f] &= \int \frac{\Delta^2(f, Q)}{\Delta^2(Q)} \cdot \frac{\Delta^2(Q)}{n!}\, \mu_n(dQ) - (f, \varphi_1)^2 \\ &= (n!)^{-1} \int [\det \| f(\boldsymbol{x}_i), \varphi_2(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i) \|]^2\, \mu_n(dQ) - (f, \varphi_1)^2 \,. \end{aligned}$$

Entsprechend dem Lemma 4.1 erhält man $\mathsf{D}\varkappa_n[f] = \delta_n - (\varphi_1, f)^2$ mit

$$\delta_n = \begin{vmatrix} (f, f) & (\varphi_2, f) & \ldots & (\varphi_n, f) \\ (f, \varphi_2) & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ (f, \varphi_n) & 0 & \ldots & 1 \end{vmatrix} \,.$$

Durch Entwicklung der Determinante δ_n nach den Elementen der letzten Zeile erhalten wir $\delta_n = \delta_{n-1} - (f, \varphi_n)^2$. Wegen $\delta_1 = (f, f)$ folgt

$$\mathsf{D}\varkappa_n[f] = (f, f) - \sum_{i=1}^{n} (f, \varphi_i)^2 = \int \left[f(x) - \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x}) \right]^2 \mu(d\boldsymbol{x}) \,,$$

womit der Satz im Fall eines regulären Systems $\{\varphi_i\}$ bewiesen ist. Ist das System $\{\varphi_i\}$ nichtregulär, so führen wir die Bezeichnung $\mathfrak{X}_n(\Delta) = \{Q : \Delta(Q) \neq 0\}$ ein. Dann gilt

$$\begin{aligned} \mathsf{D}\varkappa_n[f] &= \int \varkappa_n^2[f] \frac{\Delta^2(Q)}{n!}\, \mu_n(dQ) - (f, \varphi_1)^2 = (n!)^{-1} \int_{\mathfrak{X}_n(\Delta)} \Delta^2(f, Q)\, \mu_n(dQ) - (f, \varphi_1)^2 \\ &\leqq (n!)^{-1} \int \Delta^2(f, Q)\, \mu_n(dQ) - (f, \varphi_1)^2 = (f, f) - \sum_{i=1}^{n} (f, \varphi_i)^2 \,, \end{aligned}$$

was den Satz vollständig beweist.

Bemerkung. Werden die φ_i nicht als orthonormiert vorausgesetzt, so bleibt der Satz 4.2 gültig. Man muß dann in (4.2.14) die φ_i durch Funktionen ψ_i ersetzen, die man aus den φ_i durch Orthonormierung erhalten kann. Diese Behauptung wird durch dieselben Überlegungen bewiesen, wie sie beim Beweis von Satz 4.1 durchgeführt wurden.

Es sei jetzt $f(\boldsymbol{x}) = \mathsf{M}_\omega \zeta(\boldsymbol{x}, \omega)$ mit einer Hilbertschen Zufallsfunktion $\zeta(\boldsymbol{x}, \omega)$, wobei $\omega \in \Omega$ und $\{\Omega, B, \lambda\}$ ein Wahrscheinlichkeitsraum ist. Wir wollen nun das

Integral $J = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$ mit Hilfe einer Quadraturformel schätzen, die durch Realisierungen der zufälligen Funktion ζ für von uns gewählte $\boldsymbol{x}$-Werte gebildet wird. Wie schon früher bezeichnet $Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ die Gesamtheit der gewählten $\boldsymbol{x}$-Werte. Wir unterscheiden zwei Fälle.

Fall A. Für $\boldsymbol{x} = \boldsymbol{x}_i$ werden Realisierungen $\zeta_i = \zeta(\boldsymbol{x}_i, \omega_i)$, $i = 1, \ldots, n$, erzeugt. Dabei sind die ω_i unabhängige Realisierungen der Zufallsgröße ω. Die Größe $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$ gehört dem Stichprobenraum $\{\Omega_n, B_n, \lambda_n\}$ an.

Fall B. Für $\boldsymbol{x} = \boldsymbol{x}_i$ werden die Realisierungen ζ bei ein und derselben Realisierung ω erzeugt, d. h., man erhält $\zeta_i' = \zeta(\boldsymbol{x}_i, \omega)$, $i = 1, \ldots, n$ (ausführlicher darüber siehe Kap. V, § 1).

Wir bilden nun die Näherungssumme

$$\varkappa_n[\zeta] = \sum_{i=1}^{n} A_i \zeta_i \qquad \text{oder} \qquad \varkappa_n[\zeta'] = \sum_{i=1}^{n} A_i \zeta_i'$$

und werden, wie schon früher, Q zufällig auswählen entsprechend der Wahrscheinlichkeitsdichte $w_0(Q) = c\Delta^2(Q)$ im Verhältnis zu μ_n. Dann gelten die folgenden Sätze (Ermakow [3], Granowski und Ermakow [1]).

Satz 4.3. *Es sei* $\{\varphi_i\}$, $i = 1, \ldots, n$, *ein System linear unabhängiger Funktionen (auf dem Träger des Maßes* μ), φ_1 *sei normiert und orthogonal zu den* φ_i, $i = 2, \ldots, n$. *Wird* $\varkappa_n[\zeta]$ *durch die Gleichung* $\varkappa_n[\zeta] = \Delta(\zeta; Q)/\Delta(Q)$ *definiert und ist* Q *mit der Wahrscheinlichkeitsdichte* $w_0(Q)$ *im Verhältnis zu* μ_n *verteilt, so gelten die Gleichungen*

$$\mathsf{M}_{\omega, Q} \varkappa_n[\zeta'] = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

und

$$\mathsf{M}_{\omega, Q} \varkappa_n[\zeta] = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})\,. \tag{4.2.15}$$

Beweis. Der Beweis folgt aus der Meßbarkeit von $\varkappa_n[\zeta']$ bzw. $\varkappa_n[\zeta]$ bezüglich der Maße λ und μ bzw. λ_n und μ_n und dem Satz von Fubini.
Komplizierter ist der Beweis von

Satz 4.4. *Wenn die* φ_i, $i = 1, \ldots, n$, *ein orthonormiertes Funktionensystem bilden, dann gilt unter den Voraussetzungen von Satz* 4.3 *im Fall* A

$$\mathsf{D}_{\omega, Q} \varkappa_n[\zeta] \leqq \int \left[f(\boldsymbol{x}) - \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x})\right]^2 \mu(d\boldsymbol{x}) + \int \sigma^2(\boldsymbol{x})\, \mu(d\boldsymbol{x})\,,$$

im Fall B

$$\mathsf{D}_{\omega, Q} \varkappa_n[\zeta'] \leqq \int \left[f(\boldsymbol{x}) - \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x})\right]^2 \mu(d\boldsymbol{x}) + \int \sigma^2(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$
$$- \sum_{i=2}^{n} \int\int R(\boldsymbol{x}_1, \boldsymbol{x}_2)\, \varphi_i(\boldsymbol{x}_1)\, \varphi_i(\boldsymbol{x}_2)\, \mu(d\boldsymbol{x}_1)\, \mu(d\boldsymbol{x}_2)\,. \tag{4.2.16}$$

Dabei ist $\sigma^2(\boldsymbol{x}) = \mathsf{M}_\omega[\zeta(\boldsymbol{x}, \omega) - f(\boldsymbol{x})]^2$ *und* $R(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *die Korrelationsfunktion von* $\zeta(\boldsymbol{x}, \omega)$. *Für ein reguläres System* $\{\varphi_i\}$ *gilt in den genannten Ungleichungen das Gleichheitszeichen.*

Beweis. Wir betrachten den Fall eines regulären Systems $\{\varphi_i\}$. Der nichtreguläre Fall wird genauso wie beim Beweis von Satz 4.2 behandelt.

Fall A. Es ist

$$\mathsf{D}\varkappa_n[\zeta] = \int \lambda_n(d\boldsymbol{\omega}) \int \mu_n(dQ)\, \frac{\Delta^2(\zeta, Q)}{\Delta^2(Q)} \cdot \frac{\Delta^2(Q)}{n!} - (\varphi_1, f)^2$$

und

$$\mathsf{D}\varkappa_n[\zeta] = (n!)^{-1} \int \lambda_n(d\boldsymbol{\omega}) \int \mu_n\,(dQ)\,[\det \|\zeta_i, \varphi_2(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i)\|]^2 - (\varphi_1, f)^2 .$$

Durch Entwicklung der Determinante unter dem Integral nach den Elementen der ersten Spalte und Änderung der Numerierung der Veränderlichen ergibt sich

$$\begin{aligned}\mathsf{D}\varkappa_n[\zeta] &= \int \lambda_n(d\boldsymbol{\omega}) \int \mu_n(dQ)\,\big((n-1)!\big)^{-1}\, \zeta_1^2 \Delta_1^2(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\\ &\quad - \int \lambda_n(d\boldsymbol{\omega}) \int \mu(dQ)\,\big((n-2)!\big)^{-1}\, \zeta_1\zeta_2 \Delta_1(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\, \Delta_2(\boldsymbol{x}_1, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_n) .\end{aligned}$$

Dabei ist $\Delta_i(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_n) = \Delta_i$ der Minor des i-ten Elements der ersten Spalte der Determinante $\Delta(\zeta, Q)$.

Mit Hilfe der Resultate von Lemma 4.1 erhalten wir

$$\mathsf{D}\varkappa_n[\zeta] = \int \lambda(d\omega) \int \mu(d\boldsymbol{x})\, \zeta^2(\boldsymbol{x}, \omega) - (\varphi_1, f)^2 - \big((n-2)!\big)^{-1} J_2$$

mit

$$J_2 = \int \lambda_n(d\boldsymbol{\omega}) \int \mu_n(dQ)\, \zeta_1\zeta_2 \Delta_1 \Delta_2 .$$

Durch Entwicklung der Determinante Δ_1 nach den Elementen der Zeile, die von $\boldsymbol{x}_2$ abhängt, und der Determinante Δ_2 nach Elementen der Zeile, die von $\boldsymbol{x}_1$ abhängt, ergibt sich

$$J_2 = \int \lambda_n(d\omega) \int \mu_n(dQ)\, \zeta(\boldsymbol{x}_1, \omega_1)\, \zeta(\boldsymbol{x}_2, \omega_2) \sum_{i=2}^{n} \sum_{j=2}^{n} (-1)^{i+j}\, \Delta_{1,i} \Delta_{2,j}\, \varphi_i(\boldsymbol{x}_1)\, \varphi_j(\boldsymbol{x}_2)$$

mit den entsprechenden Minoren $\Delta_{1,i}$ und $\Delta_{2,i}$. Dabei ist $\Delta_{1,i} = \Delta_{2,i}$, $i = 1, \ldots, n-1$, und $\Delta_{1,i}$ hängt nicht von $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ ab. Aus Lemma 4.1 folgt leicht

$$\int \mu(d\boldsymbol{x}_3) \cdots \int \mu(d\boldsymbol{x}_n)\, \Delta_{1,i} \Delta_{1,j} = \begin{cases} (n-2)! & \text{für} \quad i = j ,\\ 0 & \text{sonst.} \end{cases}$$

Daher ist

$$J_2 = (n-2)! \int \lambda(d\omega_1) \int \lambda(d\omega_2) \int \mu(d\boldsymbol{x}_1) \int \mu(d\boldsymbol{x}_2)\, \zeta(\boldsymbol{x}_1, \omega_1)\, \zeta(\boldsymbol{x}_2, \omega_2) \sum_{i=2}^{n} \varphi_i(\boldsymbol{x}_1)\, \varphi_i(\boldsymbol{x}_2) . \tag{4.2.17}$$

Nun genügt es, die Integration nach ω_1 und ω_2 auszuführen, um das geforderte Resultat im Fall A zu erhalten.

Es ist leicht einzusehen, daß der Fall B analog untersucht werden kann, aber anstelle von (4.2.17) gelangen wir zu

$$J_2 = (n-2)! \int \lambda(d\omega) \int \mu(d\boldsymbol{x}_1) \int \mu(d\boldsymbol{x}_2)\, \zeta(\boldsymbol{x}_1, \omega)\, \zeta(\boldsymbol{x}_2, \omega) \sum_{i=2}^{n} \varphi_i(\boldsymbol{x}_1)\, \varphi_i(\boldsymbol{x}_2) ,$$

woraus nach Lemma 4.1 unmittelbar (4.2.16) folgt. Damit ist der Satz bewiesen.

Die Sätze 4.3 und 4.4 erlauben insbesondere die Anwendung der von uns konstruierten Quadraturformeln mit stochastischen Stützstellen für die Integration von Funktionen, deren Werte mit der Monte-Carlo-Methode berechnet werden.

Wir gehen nun auf einige Sonderfälle und Beispiele ein. Als Beispiel für ein reguläres System, das sich auf das Lebesguesche Maß auf dem Intervall [0, 1] bezieht, dient das System der Legendreschen Polynome, die auf [0, 1] orthonormiert sind. Dabei ist $\varphi_i(x) = P_{i-1}(x)$ vom Grad $i - 1$. Die Determinante $\Delta(Q)$ ist in diesem Fall bis auf einen Faktor die Vandermondesche Determinante

$$\Delta(Q) = c \det \|x_i^k\| , \qquad i = 1, \ldots, n;\; k = 0, \ldots, n-1 ,$$

die bekanntlich gleich

$$\Delta(Q) = c \prod_{i<j} (x_i - x_j) , \qquad i, j = 1, \ldots, n-1 , \tag{4.2.18}$$

ist. Aus (4.2.18) folgt unmittelbar, daß das μ_n-Maß der Menge $\{Q : \Delta(Q) = 0\}$ gleich Null ist.

Weitere Beispiele regulärer Systeme bezüglich des Lebesgueschen Maßes in einem Gebiet $\mathcal{D}$ des s-dimensionalen euklidischen Raumes sind Systeme von Polynomen in mehreren Veränderlichen und Systeme trigonometrischer Polynome in einer und mehreren Veränderlichen. Der Leser kann selbst ohne Mühe weitere ähnliche Beispiele finden. Beispiele nichtregulärer Systeme im euklidischen Raum in bezug auf ein gegebenes Wahrscheinlichkeitsmaß μ sind Systeme stückweise konstanter Funktionen. Nichtregulär sind insbesondere die Haarschen Funktionensysteme, die vielfältige Anwendung in der numerischen Mathematik finden (vgl. Sobol [7]). Sie werden folgendermaßen definiert.

Wir zerlegen $\mathfrak{X}$ in 2^m durchschnittsfremde Teilmengen $d(m, i_m)$, $i_m = 1, \ldots, 2^m$, gleichen μ-Maßes. (Wir nehmen an, daß $\mu(\mathfrak{X}) = 1$ ist.) Die Teilmenge $d(s, i_s)$, $1 \leqq s \leqq m$, $i_s = 1, \ldots, 2^{s-1}$, definieren wir durch die Gleichung

$$d(s, i_s) = d(s+1, 2i_s - 1) \cup d(s+1, 2i_s) .$$

Dann setzen wir

$$\chi_{i_s}^s = \begin{cases} 2^{s/2} & \text{für} \quad \boldsymbol{x} \in d(s+1, 2i_s - 1) , \\ -2^{s/2} & \text{für} \quad \boldsymbol{x} \in d(s+1, 2i_s) , \\ 0 & \text{sonst} , \end{cases} \tag{4.2.19}$$

$$\varphi_1(\boldsymbol{x}) \equiv 1 , \quad \varphi_j(\boldsymbol{x}) = \chi_{i_s}^s , \quad j = 2^s + i_s , \quad s = 1, \ldots, m .$$

Man prüft leicht nach, daß die Funktionen des Systems $\{\varphi_i\}$ orthogonal und normiert sind. Ist μ das Lebesguesche Maß, so sind die Haarschen Funktionen leicht zu konstruieren. Die Gestalt der ersten Haarschen Funktionen über dem Intervall [0, 1] ist in Abb. 4.1 angegeben. In diesem Fall wird das Intervall [0, 1] in gleiche Teile zerlegt. $\Delta(Q)$ wird Null, wenn zwei Punkte $\boldsymbol{x}_i$ und $\boldsymbol{x}_j$, $i \neq j$, in ein und dieselbe Menge $d(m, i_m)$ fallen. Man kann auch leicht ein Analogon der Haarschen Funktionen für den Fall angeben, daß die Mengen $d(m, i_m)$ verschiedenes Maß haben.

Interessant ist die Feststellung, daß in bezug auf ein diskretes Maß jedes beliebige Funktionensystem für $n \geqq 2$ nichtregulär ist. In diesem Fall erhält man nämlich jede Funktion durch Angabe ihrer Werte in diskreten Punkten, deren jeder ein von Null verschiedenes Maß besitzt. Dann kann man leicht zwei gleichzeitig von Null verschiedene Konstanten c_1 und c_2 wählen, so daß $c_1\varphi_1 + c_2\varphi_2 = 0$ im gege-

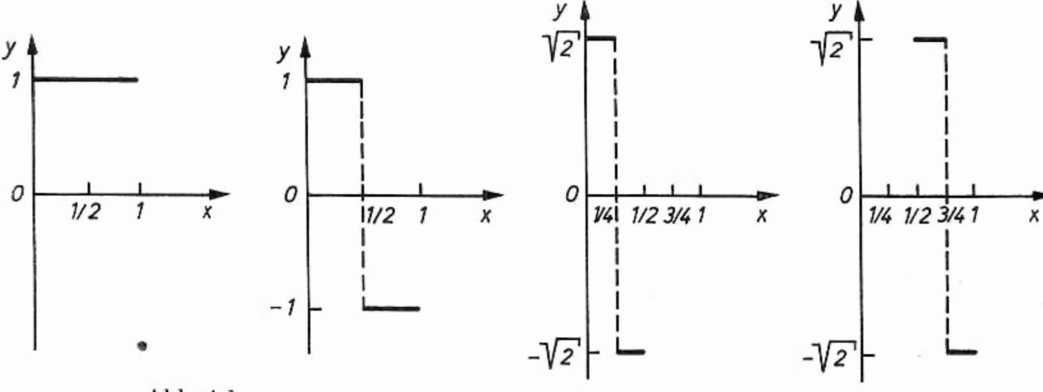


Abb. 4.1

benen Punkt gilt. Der Punkt ist aber eine Teilmenge von $\mathfrak{X}$ mit einem von Null verschiedenen Maß. Das beweist die Nichtregularität eines beliebigen Funktionensystems in bezug auf ein diskretes Maß.

Wir betrachten nun einige einfache Beispiele konkreter Quadraturformeln.

Beispiel 1. Es sei $\mathfrak{X} = [0, 1]$, μ das Lebesguesche Maß, $n = 2$, $\varphi_1(x) \equiv 1$ und $\varphi_2(x) = 2\sqrt{3}\,(x - 1/2)$. Dann ist

$$\varkappa_2[f] = f(x_1)\,\frac{x_2 - 1/2}{x_2 - x_1} + f(x_2)\,\frac{x_1 - 1/2}{x_1 - x_2}. \tag{4.2.20}$$

Die Dichte w_0 der gemeinsamen Verteilung von x_1 und x_2 ist $w_0(x_1, x_2) = 6(x_1 - x_2)^2$, d. h., die Wahrscheinlichkeit, ein Paar x_1, x_2 zu erhalten, ist um so größer, je größer das Quadrat des Abstandes zwischen x_1 und x_2 ist.

Die Dispersion von $\varkappa_2[f]$ ergibt sich aus der Gleichung

$$\mathbf{D}\varkappa_2[f] = \int_0^1 f^2(x)\,dx - \left[\int_0^1 f(x)\,dx\right]^2 - 12\left[\int_0^1 f(x)\,(x - \tfrac{1}{2})\,dx\right]^2.$$

Für die gewöhnliche Monte-Carlo-Methode beträgt die Dispersion eines Versuches unter den angegebenen Bedingungen

$$\int_0^1 f^2(x)\,dx - \left[\int_0^1 f(x)\,dx\right]^2.$$

Sie ist also größer für alle Funktionen f, die nicht zu $(x - 1/2)$ orthogonal sind. Die Benutzung von $\varkappa_2[f]$ erfordert allerdings mehr als den doppelten Aufwand für einen Versuch im Vergleich zur gewöhnlichen Monte-Carlo-Methode. Ob die Anwendung der Quadraturformel $\varkappa_2[f]$ von Vorteil ist, hängt daher von der Größe des Integrals $\int_0^1 f(x)\,(x - 1/2)\,dx$ ab. Man kann hier also nur dann eine Antwort geben, wenn zusätzliche Informationen über die Funktion f vorliegen. Eine analoge Situation liegt auch in anderen Fällen regulärer Systeme vor.

Anders verhält es sich bei nichtregulären Systemen.

Beispiel 2. Das Intervall $\mathfrak{X}$ und das Maß μ sind wie im Beispiel 1 gewählt, aber $\varphi_1(x)$ und $\varphi_2(x)$ sind zwei Haarsche Funktionen. Dann ist

$$\varkappa_2[f] = (1/2)\,[f(x_1) + f(x_2)]$$

und

$$w_0(x_1, x_2) = \begin{cases} 2 & \text{für} \quad (x_1, x_2) \in S\,, \\ 0 & \text{sonst}\,. \end{cases}$$

Dabei ist S die Punktmenge des Einheitsquadrates, die in Abb. 4.2 schraffiert worden ist (gleichgültig, ob der Rand dazugehört oder nicht). Daher sind die Randverteilungen von x_1 und x_2 Gleichverteilungen, aber x_1 und x_2 müssen jeweils in verschiedenen Hälften des Intervalls [0, 1] liegen. Es ist leicht zu sehen, daß wir auf diese Weise zu dem früher betrachteten Fall der geschichteten Stichprobe gelangen. Die Dispersion der enthaltenen Näherungssumme kann auch leicht durch die Haarschen Funktionen ausgedrückt werden:

$$\mathsf{D}\varkappa_2[f] = \tfrac{1}{2}\left\{\int_0^1 f^2(x)\,dx - \left[\int_0^1 f(x)\,dx\right]^2 - \left[\int_0^1 f(x)\,\varphi_2(x)\,dx\right]^2\right\}. \tag{4.2.21}$$

Dieser Ausdruck illustriert besonders anschaulich, daß die Schichtung vorteilhaft ist, welche Funktion aus $\mathbf{L}_2(\mu)$ auch immer zu integrieren ist. Offensichtlich führen für $n = 2^m$ die Interpolationsquadraturformeln, die für die ersten 2^m Haarschen Funktionen exakt sind, auf den Spezialfall einer geschichteten Stichprobe, wenn $\mathfrak{X}$ in 2^m Teilmengen gleichen Maßes zerlegt wird und alle n_i einander gleich sind.

Es folgen nun noch einige ergänzende Bemerkungen über Interpolationsquadraturformeln mit stochastischen Stützstellen. Wir erhielten bisher einige Aussagen über die zweckmäßige Wahl der Stützstellen einer Interpolationsquadraturformel (in Form der Verteilungsfunktion der Stützstellen) bei allgemeinsten Voraussetzungen über $f(\boldsymbol{x})$, das System $\varphi_i(\boldsymbol{x})$ und das Maß μ vom theoretischen Standpunkt. Das bewiesene Lemma und die Sätze gestatten auch eine einfache Verallgemeinerung auf den Fall komplexer Funktionen f und φ_i, $i = 1, \ldots, n$. Eine solche Verallgemeinerung wurde von Handscomb [1] betrachtet. Dort werden auch numerische Beispiele ausgeführt. Die Frage, ob W_0 die einzige Verteilungsfunktion der Stützstellen ist, die die Erwartungstreue sichert, oder in einem anderen Sinn die beste ist,

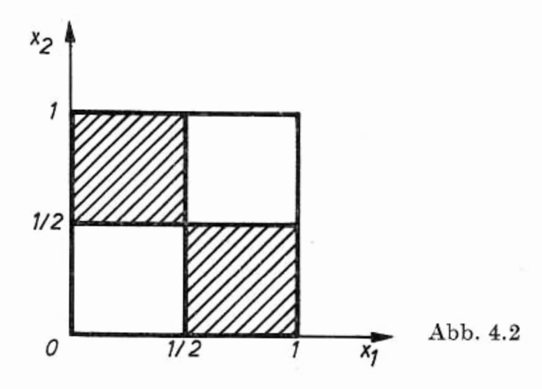


Abb. 4.2

wird in den folgenden Paragraphen untersucht. Es besteht aber kein Zweifel, daß die Resultate für die Integralberechnung nützlich sein können, wenn wir über zusätzliche Informationen über die Funktion f verfügen.

Die Modellierung einer Zufallsgröße Q mit der Dichte w_0 ist mit gewissen Schwierigkeiten verbunden. In jedem konkreten Fall kann man anscheinend die Prozedur vereinfachen, aber im allgemeinen Fall ist es schwer, irgend etwas außer der v.-Neumannschen Methode vorzuschlagen. Für großes n entsteht eine erste Schwierigkeit bei der Schätzung des Maximums des Betrages der Determinante, eine zweite bei der Berechnung der Determinante. Für nicht zu großes n (in Abhängigkeit vom Charakter der Funktionen $\varphi_i(\boldsymbol{x})$ ungefähr $n \leqq 15$) kann man hoffen, einen effektiven Algorithmus ohne ausführliche spezielle Untersuchungen zu finden. Im folgenden Kapitel kommen wir auf diese Frage im Zusammenhang mit der Interpolation zurück. Wichtig ist auch, daß die Funktionen φ_i willkürlich sind. Damit können verschiedenartige Informationen über die Funktion f berücksichtigt werden. Der Bearbeiter verfügt also immer über solche Informationen wie: die Funktion f ist „ähnlich" einer gewissen einfachen Funktion $\varphi_i(\boldsymbol{x})$.

Es scheint nützlich zu sein, noch zwei anschauliche Beispiele zu bringen, die mit der Berechnung konkreter Integrale zusammenhängen.

Beispiel 1 (Handscomb [1]). Für die Berechnung des Integrals

$$J = \int_0^1 \cdots \int_0^1 [\exp (x_1 \cdots x_{20}) - 1] \, dx_1 \cdots dx_{20} \approx 9{,}538 \cdot 10^{-7}$$

mit einer Genauigkeit von 1% durch die gewöhnliche Monte-Carlo-Methode sind etwa $3 \cdot 10^9$ Versuche erforderlich (die Dispersion eines Versuches beträgt $3 \cdot 10^{-10}$). Setzt man $n = 1$, $\varphi_1 \equiv 1$, $\varphi_2 = x_1 \cdots x_{20}$, so bietet die Modellierung der Dichte $\Delta(Q)$ keine große Schwierigkeit. Die Dispersion der so erhaltenen Näherungssumme beträgt nur noch $3 \cdot 10^{-15}$. Das bedeutet, daß zum Erreichen einer Genauigkeit von 1% nur etwa 30 Versuche erforderlich sind. Natürlich ist das ein Beispiel für eine sehr günstige Auswahl der Funktionen φ_1 und φ_2. Aber es veranschaulicht in gewisser Weise die Möglichkeiten dieser Methode.

Beispiel 2 (Ermakow [1]). Zu berechnen ist

$$J = \int_0^1 f(x) \sin (2k\pi x) \, dx$$

mit einer festen ganzen Zahl k. Wir setzen $\varphi_1(x) = 1$ und $\varphi_2(x) = \operatorname{sign}\big(\sin (2k\pi x)\big)$. Das System φ_1 und φ_2 ist nichtregulär, und man prüft leicht nach, daß es sich um ein System Haarscher Funktionen im Sinn der Definition (4.2.19) handelt. Das Intervall $[0, 1]$ wird in zwei durchschnittfremde Teilmengen zerlegt. Die erste Teilmenge enthält diejenigen x, für die $\sin (2k\pi x) \geqq 0$ ist, die zweite Teilmenge diejenigen x mit $\sin (2k\pi x) < 0$. Die Determinante $\Delta(x_1, x_2)$ verschwindet, wenn $\sin (2k\pi x_1)$ und $\sin (2k\pi x_2)$ gleiches Vorzeichen haben, und ist konstant, $\Delta(x_1, x_2) = 2$, wenn $\sin (2k\pi x_1)$ und $\sin (2k\pi x_2)$ verschiedenes Vorzeichen haben. Die Näherungssumme $\varkappa_2[f]$ ist sehr einfach

$$\varkappa_2[f] = \tfrac{1}{2} \, [f(x_1) \sin (2k\pi x_1) + f(x_2) \sin (2k\pi x_2)] \, ,$$

wobei x_1 und x_2 folgendermaßen modelliert werden: Man wählt α_1 und α_2. Ist $\varphi_2(\alpha_1)\,\varphi_2(\alpha_2) = 1$, so setzt man $x_1 = \alpha_1$, $x_2 = \alpha_2$. Andernfalls wählt man von neuem α_1 und α_2 solange, bis die angegebene Gleichung erfüllt ist. Die Dispersion von $\varkappa_2$ beträgt

$$\mathsf{D}\varkappa_2[f] = \tfrac{1}{2}\left[\int_0^1 f^2(x)\sin^2(2k\pi x)\,dx - \left(\int_0^1 f(x)\sin(2k\pi x)\,dx\right)^2 - \left(\int_0^1 f(x)\,|\sin(2k\pi x)|\,dx\right)^2\right].$$

Die gewöhnliche Monte-Carlo-Methode ergibt bei der Berechnung von zwei Funktionswerten

$$\tilde{\varkappa} = \tfrac{1}{2}\,[f(\alpha_1)\sin(2k\pi\alpha_1) + f(\alpha_2)\sin(2k\pi\alpha_2)]\,,$$

$$\mathsf{D}\tilde{\varkappa} = \frac{1}{2}\left[\int_0^1 f^2(x)\sin^2(2k\pi x)\,dx - \left(\int_0^1 f(x)\sin(2k\pi x)\,dx\right)^2\right].$$

Für nichtnegative Funktionen f erhalten wir daher einen gewissen Vorteil. Dieser kann in einzelnen Fällen bedeutend sein. So ergibt sich für $f(x) = x$

$$\mathsf{D}\tilde{\varkappa} = \frac{1}{12} - \frac{5}{32\pi^2 k^2}, \qquad \mathsf{D}\varkappa_2 = \mathsf{D}\tilde{\varkappa} - \frac{1}{2\pi^2},$$

und für beliebiges $k \geqq 1$ ist das Verhältnis der Dispersionen $\mathsf{D}\tilde{\varkappa}/\mathsf{D}\varkappa_2 \geqq 2$. Die Prüfung der Gleichung $\varphi_2(\alpha_1)\,\varphi_2(\alpha_2) = 1$ kann offensichtlich durch die Prüfung ersetzt werden, ob die größten ganzen Bestandteile der Zahlen $2k\alpha_1$ bzw. $2k\alpha_2$ beide gerade oder beide ungerade sind.

§ 3. Die allgemeine Aufgabenstellung und der Begriff der Zulässigkeit

In § 2 haben wir verschiedene erwartungstreue Schätzungen für ein Integral betrachtet. Vom Standpunkt der mathematischen Statistik sind die Methode der Abtrennung des Hauptteils und die Methode der Erniedrigung der Integrationsordnung Verfahren, die direkt erwartungstreue Schätzungen verschiedener Art benutzen. Der Vorteil jener Aufgaben, die in der Monte-Carlo-Methode entstehen, gegenüber den Aufgaben der Konstruktion statistischer Schätzungen besteht in der Möglichkeit, über die Wahrscheinlichkeitsverteilungen zu verfügen. Das kommt anschaulich in der Methode der wesentlichen Stichprobe und bei der Anwendung von Interpolationsquadraturformeln zum Ausdruck.

In diesem Paragraphen werden wir die Untersuchungen von stochastischen Quadraturformeln fortsetzen. Vor allem ergibt sich die natürliche Frage, ob W_0 die einzige Verteilungsfunktion der Stützstellen einer Interpolationsnäherungssumme ist, die ihre Erwartungstreue sichert. Es ist leicht zu sehen, daß das nicht so ist. Bekannt ist beispielsweise das Resultat von HAMMERSLEY und MORTON [1] (antithetic variate method, oder wie wir sagen werden, die Methode der sym-

metrischen Stichprobe). In der einfachsten Form besteht sie in der Anwendung einer stochastischen Quadraturformel für die Berechnung des Integrals über die Funktion $f(x)$ auf dem Intervall $[0, 1]$ von der Gestalt

$$\int_0^1 f(x)\,dx \approx \tfrac{1}{2}\,[f(x) + f(1-x)]\,, \tag{4.3.1}$$

wobei x auf $[0, 1]$ gleichverteilt ist.

Die Formel (4.3.1) war die erste stochastische Quadraturformel, die in der Literatur betrachtet wurde. Ihr Studium trug wesentlich zur Entwicklung von Arbeiten über die Dispersionsverkleinerung bei der Monte-Carlo-Methode bei. Die Formel (4.3.1) gilt exakt für 1 und eine beliebige bezüglich der Geraden $x = 1/2$ ungerade Funktion und ist wesentlich einfacher als etwa die Formel (4.2.13), die nur für 1 und x exakt gilt. Offensichtlich ist auch ihre Erwartungstreue. Im Unterschied zu den Formeln, die in § 2 betrachtet wurden, wird in (4.3.1) nur eine Stützstelle stochastisch gewählt, während die zweite durch diese Wahl eindeutig mitbestimmt ist. (Die Verteilungsfunktion der Stützstellen ist ausgeartet.) Ausführlich werden wir Formeln solcher Art in § 4 untersuchen. Jetzt gehen wir auf die allgemeinen Bedingungen ein, die eine Verteilungsfunktion der Stützstellen erfüllen muß, damit die Erwartungstreue der Näherungssumme (die nicht notwendig vom Interpolationstyp sein muß) gesichert ist.

Wir werden Summen der Gestalt

$$\varkappa_n[f] = \sum_{i=1}^{n} A_i(Q)\, f(\boldsymbol{x}_i)$$

betrachten, um das Integral

$$J = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

zu berechnen. Dabei ist $Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ eine mit der Gesetzmäßigkeit $U(Q)$ verteilte Zufallsgröße, und die $A_i(Q)$ sind gewisse gegebene Funktionen. Neben der Bezeichnung $\varkappa_n[f]$ werden wir auch die Bezeichnungen

$$\varkappa_n[f] = \varkappa_n[f; Q] = \varkappa_n[f; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$$

gebrauchen. Als Kriterium für die Güte einer erwartungstreuen Summe in $\mathbf{L}_2(\mu)$ werden wir ihre Dispersion benutzen, die als endlich vorausgesetzt wird. Dabei beschränken wir uns auf die Betrachtung von Summen, die symmetrisch bezüglich der Variablen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ sind und für die $U(Q)$ auch eine symmetrische Funktion der genannten Variablen ist. Dies bedeutet vom Standpunkt des gewählten Kriteriums her keine Einschränkung der Allgemeinheit.

Wenn nämlich $\tilde{\varkappa}_n$ und $\tilde{U}(Q)$ die genannte Symmetrie nicht besitzen und die Summe $\tilde{\varkappa}_n$ erwartungstreu ist, dann konstruieren wir folgendermaßen ein symmetrisches $\varkappa_n$ und $U(Q)$. Wir setzen

$$U(Q) = \frac{1}{n!} \sum_{1 \leqq i_1 < \cdots < i_n \leqq n} \tilde{U}(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_n}) \tag{4.3.2}$$

und

$$\varkappa_n = \frac{1}{n!} \sum \tilde{\varkappa}_n\,[f; \boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_n}]\, \frac{d\tilde{U}}{dU}\,(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_n})\,.$$

(Man prüft leicht nach, daß $\tilde{U}$ bezüglich U absolut stetig ist.) Wenn wir der Einfachheit halber die Bezeichnung der Variablen und die Summationsindizes weglassen, dann ergibt sich

$$\mathsf{M}\varkappa_n = \int \frac{1}{n!} \Sigma \, \tilde{\varkappa}_n \frac{d\tilde{U}}{dU} \, dU = \frac{1}{n!} \Sigma \int \tilde{\varkappa}_n \, d\tilde{U} \, ,$$

d. h., $\varkappa_n$ ist eine erwartungstreue Schätzung.

Weiter gilt $\mathcal{D}\varkappa_n = \mathsf{M}\varkappa_n^2[f] - J^2$ und

$$\varkappa_n^2[f] = \frac{1}{(n!)^2} \left(\Sigma \, \tilde{\varkappa}_n \frac{d\tilde{U}}{dU} \right)^2 \leqq \frac{1}{(n!)^2} \Sigma \, \tilde{\varkappa}_n^2[f] \frac{d\tilde{U}}{dU} \, \Sigma \frac{d\tilde{U}}{dU} = \frac{1}{n!} \Sigma \, \tilde{\varkappa}_n^2[f] \frac{d\tilde{U}}{dU}$$

wegen

$$\Sigma \frac{d\tilde{U}}{dU} = \frac{\Sigma \, d\tilde{U}}{dU} = n! \, .$$

Daraus erhalten wir

$$\mathsf{D}\varkappa_n[f] \leqq \mathsf{D}\tilde{\varkappa}_n[f]$$

für alle $f \in \mathbf{L}_2(\mu)$. Daher ist die symmetrisierte Näherungssumme vom Standpunkt der Größe der Dispersion nicht schlechter als die ursprüngliche Summe.

Eine solche Gesamtheit $(\varkappa_n[f], U(Q))$, für die $\varkappa_n[f]$ für alle $f \in \mathbf{F}$ erwartungstreu ist und $\mathsf{D}\varkappa_n[f] < \infty$ für $f \in \mathbf{F}$ gilt, werden wir im weiteren eine Prozedur der Monte-Carlo-Methode (PMCM) im engeren Sinne für die Funktionenklasse $\mathbf{F}$ nennen. Damit $(\varkappa_n[f], U(Q))$ eine PMCM (die Bezeichnung „im engeren Sinne" lassen wir im folgenden weg) ist, müssen folgende Bedingungen erfüllt sein:

Bedingung A (Erwartungstreue). Es muß

$$\int \varkappa_n[f] \, U(dQ) = \int \varphi_1(\boldsymbol{x}) \, f(\boldsymbol{x}) \, \mu(d\boldsymbol{x}) \tag{4.3.3}$$

für beliebiges $f \in \mathbf{F}$ gelten.

Wenn U in bezug auf das Maß μ eine Dichte u besitzt und $\mathbf{F}$ eine genügend umfangreiche Funktionenklasse ist, etwa $\mathbf{F} = \mathbf{L}_2(\mu)$, dann kann (4.3.3) in der Form

$$\int \mu(d\boldsymbol{x}_1) \cdots \int \mu \, (d\boldsymbol{x}_{i-1}) \int \mu(d\boldsymbol{x}_{i+1}) \cdots \int \mu \, (d\boldsymbol{x}_n) \, A_i(Q) \, u(Q) = n \, \varphi_1(\boldsymbol{x}_i) \tag{4.3.4}$$

für fast alle $\boldsymbol{x}$ geschrieben werden.

Bedingung B (Endlichkeit der Dispersion). Es muß gelten:

$$\mathsf{D}\varkappa_n[f] < \infty$$

für beliebiges $f \in \mathbf{F}$ gelten.

Wir werden uns nun mit dem Vergleich der Güte verschiedener Prozeduren der Monte-Carlo-Methode beschäftigen. In § 2 begegneten wir im Zusammenhang mit den Haarschen Funktionen dem Sachverhalt, daß man eine Prozedur der Monte-Carlo-Methode angeben kann, die in einem gewissen Sinne „besser" ist als

die gewöhnliche Monte-Carlo-Methode für alle Funktionen aus $\mathbf{L}_2(\mu)$. Es ist klar, daß eine Erscheinung solcher Art ein ausführlicheres Studium verdient. Für eine zufällige Näherungssumme $\varkappa_n[f]$ verwenden wir jetzt die Bezeichnung $\varkappa[f, Q]$. Die Größe n wird also als fest betrachtet, aber wir unterstreichen die Abhängigkeit von der Zufallsgröße Q. Wenn es wesentlich ist, auf das Verteilungsgesetz von Q hinzuweisen, werden wir hier und im folgenden die Bezeichnung Q_W verwenden, wobei W die Verteilungsfunktion von Q ist, oder Q_w, wenn w die W entsprechende Dichte bedeutet.

Es sei nun eine gewisse Klasse von Prozeduren $\mathcal{P} = (\mathcal{K}, \mathbf{U})$, $\varkappa \in \mathcal{K}$, $U \in \mathbf{U}$, und eine Klasse $\mathbf{F}$ von Funktionen f, $\mathbf{F} \subseteqq \mathbf{L}_2(\mu)$ gegeben.

Wenn es ein $p' = (\varkappa', U') \in \mathcal{P}$ gibt, so daß $\mathsf{D}\varkappa'[f, Q_{U'}] \leqq \mathsf{D}\varkappa[f, Q_U]$ für alle $f \in \mathbf{F}$ und $\mathsf{D}\varkappa'[\tilde{f}, Q_{U'}] < \mathsf{D}\varkappa[\tilde{f}, Q_U]$ für wenigstens ein $\tilde{f} \in \mathbf{F}$ gilt, dann sagen wir, daß p' über $p = (\varkappa, U)$ in $\mathcal{P}$ für $\mathbf{F}$ dominiert.

Wir nennen ein p unzulässig in $\mathcal{P}$ für $\mathbf{F}$, wenn es in $\mathcal{P}$ ein p' gibt, das über p für $\mathbf{F}$ dominiert. Andernfalls heißt p zulässig. Der hier eingeführte Begriff der Zulässigkeit ist dem oft in der Statistik gebrauchten Zulässigkeitsbegriff für eine Lösungsprozedur oder eine Schätzung völlig analog (vgl. etwa Lehmann [1]). Eine für $\mathbf{F}$ zulässige Prozedur kann man auch *nichtverbesserungsfähig* für die Funktionenklasse $\mathbf{F}$ in der Klasse der Prozeduren $\mathcal{P}$ nennen.

Der Begriff der Zulässigkeit bezieht sich auf eine Funktionenklasse. Für eine gewisse konkrete, genügend einfach zu berechnende Funktion f kann eine unzulässige Prozedur zu einem kleineren Rechenaufwand führen als eine zulässige (wegen der Notwendigkeit, $\varkappa[f, Q]$ zu berechnen). Es ist aber klar, daß man in einer genügend umfangreichen Funktionenklasse $\mathbf{F}$ stets Funktionen findet, für die die umgekehrte Behauptung gilt, da der Aufwand zur Berechnung der Funktionswerte häufig bestimmend sein wird.

Aus den Ergebnissen von § 2 folgt vor allem, daß die gewöhnliche Monte-Carlo-Methode für $\mathbf{L}_2(\mu)$ sogar in der Klasse von Prozeduren unzulässig ist, in der die Näherungssumme gleiche Koeffizienten hat, aber die Stützstellen nicht unbedingt unabhängige Zufallsgrößen sein müssen. Daher kann man also behaupten, daß die gewöhnliche Monte-Carlo-Methode, eine der einfachsten Integrationsmethoden, sogar für die sehr umfangreiche Funktionenklasse $\mathbf{L}_2(\mu)$ bei weitem nicht die beste ist.

Wir wählen jetzt ein festes Funktionensystem $\{\varphi_i\}$, $i = 1, \ldots, n$, $\mathbf{F} = \mathbf{L}_2(\mu)$ und nehmen an, daß $\varkappa[f]$ durch die Formel (4.1.6)

$$\varkappa[f] = \frac{\Delta(f, Q)}{\Delta(Q)}$$

definiert wird. Wir betrachten die Klasse von Prozeduren $\mathcal{P}_0 = (\varkappa[f], \mathbf{U})$, in der $\mathbf{U}$ die Menge von Wahrscheinlichkeitsmaßen auf $\mathfrak{X}$ ist, für die $\varkappa[f]$ erwartungstreu ist. Dann gilt

Satz 4.5 (Ermakow [4]). *Sind die Funktionen φ_i, $i = 1, \ldots, n$, in bezug auf das Wahrscheinlichkeitsmaß μ regulär und mit der Wahrscheinlichkeit 1 beschränkt und gibt es eine mit der Wahrscheinlichkeit 1 beschränkte Funktion $\psi(\boldsymbol{x})$, die zu allen Produkten $\varphi_i(\boldsymbol{x})\,\varphi_j(\boldsymbol{x})$ und zu allen $\varphi_i(\boldsymbol{x})$ orthogonal ist ($i, j = 1, \ldots, n$), dann ist die Prozedur $(\varkappa[f], W_0)$ unzulässig in $\mathcal{P}_0$ für $\mathbf{L}_2(\mu)$.*

Beweis. Wir betrachten die durch die Gleichung

$$V(dQ) = \left(1 + \beta \sum_{p<q} \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\right) W_0(dQ)\,, \qquad p, q = 1, \ldots, n\,,$$

definierte Verteilung $V(dQ)$, wobei $\psi(\boldsymbol{x})$ die durch den Satz bestimmte Funktion und $\beta > 0$ eine Konstante mit

$$1 + \beta \sum_{p<q} \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q) \geqq 0$$

für fast alle $\boldsymbol{x}$ in $\mathfrak{X}_n$ ist. Die Existenz eines solchen β folgt aus der Beschränktheit von $\psi(\boldsymbol{x})$. Daher ist $V(dQ)$ nichtnegativ. Den Beweis führen wir in zwei Schritten.

1. Wir zeigen, daß $V(\mathfrak{X}) = 1$ und $\mathsf{M}\varkappa[f; Q_V] = J$ ist. Entsprechend der Definition von $\varkappa[f]$ und $W_0(dQ)$ ergibt sich

$$\int \varkappa[f]\; V(dQ) = \frac{1}{n!} \int \frac{\Delta(f; Q)}{\Delta(Q)}\, \Delta^2(Q) \left(1 + \beta \sum_{p<q} \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\right) \mu_n(dQ)$$

$$= \mathsf{M}\varkappa[f; Q_{W_0}] + \frac{\beta}{n!} \sum_{p<q} \int \Delta(f, Q)\, \Delta(Q)\, \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\, \mu_n(dQ)\,.$$

Wegen $\mathsf{M}\varkappa[f; Q_{W_0}] = J$ (Satz 4.1) genügt es zu zeigen, daß

$$\int \Delta(f, Q)\, \Delta(Q)\, \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\, \mu_n(dQ) = 0\,, \quad p, q = 1, \ldots, n\,, \quad p \neq q\,, \tag{4.3.5}$$

und $V(\mathfrak{X}_n) = 1$ ist. Wir beweisen zunächst die erste dieser Behauptungen. Jede Determinante unter dem Integralzeichen von (4.3.5) stellt eine Summe von Produkten aus jeweils einem Element jeder Zeile und jeder Spalte, versehen mit dem entsprechenden Vorzeichen, dar. Das Produkt dieser Determinanten besteht aus Summanden, in die als von $\boldsymbol{x}_p$ abhängige Faktoren entweder $\varphi_{i_0}(\boldsymbol{x}_p)\, \varphi_{j_0}(\boldsymbol{x}_q)$ oder $f(\boldsymbol{x}_p)\, \varphi_{i_0}(\boldsymbol{x}_p)$ bei gewissen $i_0 \neq j_0$, $i_0, j_0 = 1, \ldots, n$, eingehen. Weil $f(\boldsymbol{x}_p)$ in jedem Summanden nur einmal (als Faktor) enthalten ist, geht in den Summanden der $(f(\boldsymbol{x}_p)\, \varphi_{i_0}(\boldsymbol{x}_p)$ enthält, notwendigerweise der Ausdruck $\varphi_{i_1}(\boldsymbol{x}_q)\, \varphi_{j_1}(\boldsymbol{x}_q)$, $i_1 \neq j_1$ $i_1, j_1 = 1, \ldots, n$, als weiterer Faktor ein, und es gibt darin keinen anderen von $\boldsymbol{x}_q$ abhängigen Faktor. Daher erhalten wir nach der Integation über das Produkt der mit $\psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)$ multiplizierten Determinanten wegen der Orthogonalität von $\psi(\boldsymbol{x})$ zu jedem Paar $\varphi_i(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})$ den Wert Null. Insbesondere ergibt sich für $f \equiv \varphi_1$

$$\varkappa[f] \equiv 1 \qquad \text{und} \qquad \int \varkappa[f]\; V(dQ) = \int V(dQ) = 1\,.$$

Also ist V ein Wahrscheinlichkeitsmaß und (4.3.5) bedeutet, daß

$$\int \varkappa[f]\; V(dQ) = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

ist, d. h. $\mathsf{M}\varkappa[f; Q_V] = J$.

2. Wir berechnen nun die Dispersion von $\varkappa[f]$. Aus der Regularität der φ_i folgt unmittelbar

$$\mathsf{D}\varkappa[f; Q_V] = \mathsf{D}\varkappa[f; Q_{W_0}] + \beta\; R/n!$$

mit

$$R = \sum_{p<q} \int \Delta^2(f; Q)\, \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\, \mu_n(dQ)\,.$$

Wir zeigen, daß $R \leqq 0$ für jedes $f \in \mathbf{L}_2(\mu)$ gilt. Es ist

$$R = \sum_{p<q} \int \sum_{i,j} f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_j)\, (-1)^{i+j}\, \Delta_i \Delta_j\, \psi(\boldsymbol{x}_p)\, \psi(\boldsymbol{x}_q)\, \mu_n(dQ)\,,$$

wobei Δ_i der Minor des i-ten Elementes der ersten Spalte von der Determinante im Integranden ist. Durch entsprechende Überlegungen wie im ersten Schritt des Beweises zeigt man leicht, daß eine Reihe von Summanden im letzten Ausdruck unter dem Integral nach der Integration Null wird. Damit ergibt sich für R der Ausdruck

$$\begin{aligned} R &= \int \sum_{i \neq j} f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_j)\, (-1)^{i+j}\, \Delta_i \Delta_j \psi(\boldsymbol{x}_i)\, \psi(\boldsymbol{x}_j)\, \mu_n(dQ) \\ &= \int \{\det \|f(\boldsymbol{x}_i)\, \psi(\boldsymbol{x}_i), \varphi_2(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i)\|\}^2\, \mu_n(dQ) \\ &\quad - \sum_{i=1}^{n} \int f^2(\boldsymbol{x}_i)\, \Delta_i^2 \psi^2(\boldsymbol{x}_i)\, \mu_n(dQ)\,. \end{aligned}$$

Unter Benutzung von Lemma 4.1 und Satz 4.2 erhalten wir

$$\begin{aligned} &\int \{\det \|f(\boldsymbol{x}_i)\, \psi(\boldsymbol{x}_i), \varphi_2(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i)\|\}^2\, \mu_n(dQ) \\ &= n!\left\{\int f^2(\boldsymbol{x})\, \psi^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) - \sum_{i=2}^{n} \left(\int f(\boldsymbol{x})\, \varphi_i(\boldsymbol{x})\, \psi(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right)^2\right\}, \\ &\int f^2(\boldsymbol{x}_i)\, \Delta_i^2 \psi^2(\boldsymbol{x}_i)\, \mu_n(dQ) = (n-1)! \int f^2(\boldsymbol{x})\, \psi^2(\boldsymbol{x})\, \mu(d\boldsymbol{x})\,. \end{aligned}$$

Schließlich ergibt sich

$$\mathsf{D}\varkappa[f; Q_V] = \mathsf{D}\varkappa[f; Q_{W_0}] - \beta \sum_{i=2}^{n} \left(\int f(\boldsymbol{x})\, \varphi_i(\boldsymbol{x})\, \psi(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right)^2.$$

Für $\tilde{f} = \psi\varphi_i$ mit $\tilde{f} \in \mathbf{L}_2(\mu)$, weil ψ mit der Wahrscheinlichkeit 1 beschränkt ist, erhalten wir $\mathsf{D}\varkappa[\tilde{f}, Q_{W_0}] > \mathsf{D}\varkappa[\tilde{f}, Q_V]$, womit der Beweis des Satzes abgeschlossen ist.

Bemerkung 1. Die Benutzung von V anstelle von W_0 ist eine Methode, um die Dispersion der stochastischen Näherungssumme zu verkleinern. Wenn es m linear unabhängige Funktionen $\psi_1, \ldots, \psi_m$ mit entsprechenden Eigenschaften wie ψ und m nichtnegative Konstanten β_i mit

$$1 + \sum_{i=1}^{m} \beta_i \sum_{p<q} \psi_i(\boldsymbol{x}_q)\, \psi_i(\boldsymbol{x}_p) \geqq 0 \qquad \text{für alle } \boldsymbol{x}_p \text{ und } \boldsymbol{x}_q$$

gibt, dann kann man mit einer analogen Methode ein $V^{(m)}$ konstruieren, so daß

$$\mathsf{D}\varkappa[f; Q_{V(m)}] = \mathsf{D}\varkappa[f; Q_{W_0}] - \sum_{i=1}^{m} \beta_i \sum_{j=2}^{n} \left(\int f\psi_i\varphi_j\, \mu(d\boldsymbol{x})\right)^2$$

gilt. Das Problem, ob so eine zulässige Prozedur konstruiert werden kann, bleibt hier offen. Es gilt aber der

Satz 4.6. *Wenn $\varkappa[f]$ durch die Gleichung* (4.1.6) *definiert wird, wobei die φ_i, $i = 1, \ldots, n$ die ersten n Haarschen Funktionen sind und $\mathcal{P}_1 = (\varkappa[f], \mathbf{U})$ eine Klasse von Prozeduren ist, für die $U \in \mathbf{U}$ eine Dichte in bezug auf das Maß μ ist, dann ist die Prozedur $(\varkappa[f], u)$ für jedes $U \in \mathbf{U}$ zulässig.*

Beweis. Der Einfachheit halber beweisen wir den Satz für den Fall $n = 2^m$, m ganze Zahl. Der allgemeine Fall (vgl. GRANOWSKI [1]) wird analog bewiesen. Die n ersten Haarschen Funktionen sind konstant auf jeder der Mengen $d(m, i_m)$, $i_m = 1, \ldots, 2^m$. Folglich sind die Determinante $\Delta(Q)$ und alle ihre Minoren konstant auf jeder der $n!$ Mengen $d_m^n(l_1, \ldots, l_n)$ der Art

$$d_m^n(l_1, \ldots, l_n) = d(m, l_1) \times \cdots \times d(m, l_n) ,$$

wobei $l_1 \neq l_2 \neq \cdots \neq l_n$ eine beliebige Permutation der Zahlen $1, \ldots, n$ ist und $\Delta(Q) = 0$ nur außerhalb dieser Mengen ist. Da $\varkappa[f; Q]$ nur für solche Q definiert ist, für die $\Delta(Q) \neq 0$ ist, ist $u(Q)$ auch nur für diese Q von Null verschieden. Die Summe $\varkappa[f]$ hat wegen ihrer Symmetrie die Gestalt

$$\varkappa[f] = n^{-1} \sum_{i=1}^{n} f(\boldsymbol{x}_i)$$

mit $\boldsymbol{x}_i \in d(m, i_m)$. Die Bedingung (4.3.3) der Erwartungstreue kann in der Form

$$n^{-1} \sum \int f(\boldsymbol{x}_i)\ u(Q)\, \mu_n(dQ) = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{4.3.6}$$

für beliebiges $f \in \mathbf{L}_2(\mu)$ geschrieben werden. Ferner gilt

$$\begin{aligned}\mathsf{M}\varkappa^2[f] &= n^{-2} \sum_{i,j} \int f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_j)\, u(Q)\, \mu_n(dQ) \\ &= n^{-2} \sum_{i=1}^{n} \int f^2(\boldsymbol{x}_i)\, u(Q)\, \mu_n(dQ) + n^{-2} \sum_{i \neq j} \int f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_j)\, u(Q)\, \mu_n(dQ)\end{aligned}$$

und infolge der Symmetrie

$$\mathsf{M}\varkappa^2[f] = n^{-1} \int f^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) + (n-1)\, n^{-1} \int f(\boldsymbol{x}_1)\, f(\boldsymbol{x}_2)\, u(Q)\, \mu_n(dQ) .$$

Wir nehmen nun an, daß die Prozedur $p' = (\varkappa, U')$ über p dominiert. Dann gilt

$$\mathsf{D}\varkappa[f; Q_u] - \mathsf{D}\varkappa[f; Q_{u'}] = \int f(\boldsymbol{x}_1)\, f(\boldsymbol{x}_2)\, [u(Q) - u'(Q)]\, \mu_n(dQ) \geqq 0 \tag{4.3.7}$$

für alle $f \in \mathbf{L}_2(\mu)$. Aber u und u' sind nur in den Gebieten $d_m^n(l_1, \ldots, l_n)$ von Null verschieden. Daraus folgt sofort die Möglichkeit, ein solches f zu wählen, daß die Ungleichung (4.3.7) verletzt wird. Damit ist der Satz bewiesen.

Folgerung. *Für den Fall Haarscher Funktionen ist die Prozedur* $(\varkappa[f], W_0)$ *zulässig in* $\mathcal{P}_1$ *für* $\mathbf{L}_2(\mu)$.

Wie schon bemerkt, kann man die Quadraturformeln mit stochastischen Stützstellen für die Integration von Zufallsfunktionen benutzen. Dafür wird der Begriff der Zulässigkeit analog eingeführt. Es sei $\xi(\boldsymbol{x}, \omega)$ eine Hilbertsche Zufallsfunktion, wobei ζ dem auf $(\Omega, \mathfrak{B}, \lambda)$ definierten Raum $\mathbf{G}$ der Hilbertschen Funktionen angehört. Wir bezeichnen mit ζ_i, $i = 1, \ldots, n$, Realisierungen von $\zeta(\boldsymbol{x}, \omega)$ für $\boldsymbol{x} = \boldsymbol{x}_i$. Ist dann $(\varkappa, U) = p$ eine Prozedur der Monte-Carlo-Methode, so ist $\mathsf{M}_\omega \mathsf{M}_Q \varkappa[\zeta] = J$, und als Maß für die Güte der Prozedur p kann man die Größe $\mathsf{D}_\omega \mathsf{D}_Q \varkappa[\zeta]$ wählen. Daher unterscheidet sich die Definition der Zulässigkeit von der bisherigen nur dadurch, daß die Funktionenklasse $\mathbf{F}$ durch die konkrete Klasse $\mathbf{G}$ der Hilbertschen Zufallsfunktionen ersetzt wird. Dann gilt

Lemma 4.2. *Notwendig und hinreichend für die Zulässigkeit der Prozedur* p *für* $\mathbf{G}$ *ist ihre Zulässigkeit für* $\mathbf{L}_2(\mu)$.

Beweis. Wegen $\mathbf{L}_2(\mu) \subset \mathbf{G}$ ist das Lemma hinreichend. Für jedes feste ω ist umgekehrt $\xi(\boldsymbol{x}) \in \mathbf{L}_2(\mu)$, woraus die Notwendigkeit folgt. Daher sind die nicht zu verbessernden Prozeduren für eine umfangreiche Klasse deterministischer Funktionen auch für die Zufallsfunktionen nicht zu verbessern. Dies weist auf die enge Beziehung der von uns betrachteten Aufgaben zu Aufgaben der Versuchsplanung hin. Für zulässige Prozeduren gibt uns U die Information über die Lage der günstigen (im Sinne nicht zu verkleinernder Dispersion) Stützstellen einer Quadraturformel. In Kapitel V wird dieser Zusammenhang ausführlicher untersucht.

§ 4. Stochastische Quadraturformeln mit einer freien Stützstelle

In § 3 wurde gezeigt, daß es unendlich viele Verteilungsfunktionen für die Stützstellen einer Näherungssumme gibt, die ihre Erwartungstreue sichern. Außerdem ist eine bestimmte Reduktion der Aufgabenstellung möglich, wenn man sich auf die zulässigen Prozeduren beschränkt. Dabei soll noch einmal unterstrichen werden, daß in der Monte-Carlo-Methode die Unzulässigkeit einer Prozedur (im Sinn des vorigen Paragraphen) noch nicht ihre Unbrauchbarkeit für die rechnerische Anwendung bedeutet. Der Rechenaufwand bei Anwendung einer unzulässigen Prozedur kann kleiner sein als bei Anwendung der zulässigen Prozedur. Ist freilich die betrachtete Funktionenklasse genügend umfangreich, dann gibt es in dieser Klasse immer solche Funktionen, daß der für die Funktionswerteberechnung notwendige Aufwand den für die Berechnung der entsprechenden Quadraturformel notwendigen Aufwand übersteigt. Dadurch erklärt sich einerseits die Notwendigkeit, zulässige Prozeduren zu studieren. Andererseits sind die zulässigen Prozeduren nicht unbedingt in allen Fällen komplizierter als die unzulässigen, wie wir im weiteren sehen werden.

Anschließend wird eine spezielle Klasse zulässiger Prozeduren der Monte-Carlo-Methode studiert, die stochastischen Quadraturformeln mit einer freien Stützstelle. Ermakow [2] schlug vor, diese Formeln als Verallgemeinerung der Methode der symmetrischen Stichprobe zu betrachten. Ihr systematisches Studium wurde durch Granowski [1] bis [3] vorgenommen.

Der allgemeine Gedankengang bei der Herleitung von Quadraturformeln mit einer freien Stützstelle als Interpolationsquadraturformel besteht in folgendem.

Es sei $\varkappa_n[f]$ die Interpolationsnäherungssumme für die Funktion $f \in \mathbf{L}_2(\mu)$:

$$\varkappa_n[f] = \frac{\Delta(f; Q)}{\Delta(Q)}, \tag{4.4.1}$$

die für die orthonormierten Funktionen $\varphi_i \in \mathbf{L}_2(\mu)$, $i = 1, \ldots, n$ exakt gilt. Folgt man den allgemeinen Ideen bei der Herleitung deterministischer Formeln vom Gaußschen Typ, so muß man erreichen, daß die Quadraturformel $J \approx K_n[f]$ auch für die Funktionen $\varphi_{n+1}, \ldots, \varphi_{n+m}$ aus dem vollständigen orthonormierten System $\{\varphi_i\}$ exakt gilt. Wie schon in § 1 bemerkt, müssen dabei die Stützstellen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ auf der Hyperfläche liegen, die durch die Gleichung

$$\Delta(\varphi_{n+r}; Q) \equiv \det \|\varphi_{n+r}(\boldsymbol{x}_i), \varphi_2(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i)\| = 0 \tag{4.4.2}$$

definiert wird. Genauso kommen wir zur Herleitung einer stochastischen Quadraturformel, deren Stützstellen auf der Hyperfläche (4.4.2) verteilt sind. Diese Aufgabe ist bedeutend schwieriger als die entsprechende deterministische Aufgabe. Das liegt daran, daß wir Beziehungen zwischen den stochastischen Stützstellen finden müssen, die nicht aus den Gleichungen (4.4.2), sondern aus der Bedingung der Erwartungstreue der sich dabei ergebenden Näherungssumme $\varkappa_n[f]$ folgen. Daher kann die stochastische Quadratur nur in dem Fall durchgeführt werden, in dem die Gleichung (4.4.2) höchstens eine freie Stützstelle besitzt.

Nun sei $(\mathfrak{X}, \mathfrak{A}, \mu)$ ein Maßraum und

$$\varkappa_n[f] = \sum_{i=1}^{n} A_i(Q)\, f(\boldsymbol{x}_i)$$

eine Näherungssumme (nicht unbedingt vom Interpolationscharakter), für die alle Stützstellen Funktionen einer Stützstelle sind, etwa der ersten

$$\boldsymbol{x}_i = T_{i-1}(\boldsymbol{x}_1)\,, \qquad i = 2, \ldots, n\,. \tag{4.4.3}$$

Die Stützstelle $\boldsymbol{x}_1$ sei dabei zufällig gewählt. Wir werden annehmen, daß die Abbildungen T_i eine Funktionaldeterminante $|\Phi_i|$ besitzen. Es soll also die Radon-Nikodymsche Ableitung des Maßes ν nach dem Maß μ mit $\nu(A) = \mu(T(A))$, $A \in \mathfrak{A}$, existieren.

Ist daher die Stützstelle $\boldsymbol{x}_1$ in $\mathfrak{X}$ mit der Dichte u in bezug auf das Maß μ verteilt, so ist die gemeinsame Verteilung der Stützstellen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ im Raum $\mathfrak{X}^{(n)} = \mathfrak{X} \times \times \cdots \times \mathfrak{X}$ (n Faktoren) ausgeartet: diese Stützstellen sind auf der Hyperfläche in $\mathfrak{X}^{(n)}$ verteilt, die durch die Beziehung T_i definiert ist (insbesondere wird das für eine Interpolationsquadraturformel mit stochastischen Stützstellen eine Hyperfläche der Gestalt (4.4.2)).

Stochastische Quadraturformeln, deren Stützstellen durch die Beziehungen (4.4.3) miteinander verknüpft sind, werden wir nun stochastische Quadraturformeln mit einer freien Stützstelle nennen.

Wenn wir uns nach den Überlegungen von § 3 richten, dann können wir uns auf die Betrachtung symmetrischer Quadraturformeln beschränken. Dazu wird angenommen, daß die von den Stützstellen geforderten Beziehungen symmetrisch sind (d. h., alle Stützstellen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ sind gleichberechtigt. Eine Symmetrie solcher Art führt uns zu folgenden Forderungen, die an die T_i zu stellen sind:

$$\begin{gathered} T_i = T^i\,, \qquad i = 0, \ldots, n-1\,, \\ \text{mit} \quad T = T_1 \qquad \text{und} \qquad T^n = T^0 \neq T^i\,, \qquad i < n\,. \end{gathered} \tag{4.4.4}$$

Wir werden also fordern, daß die Abbildungen T_i eine zyklische Gruppe der Ordnung n bilden, die durch die Abbildung T erzeugt wird. Dann gilt der folgende

Satz 4.7. *Die symmetrische stochastische Quadraturformel mit einer freien Stützstelle*

$$\varkappa_n[f] = \int \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{4.4.5}$$

soll für die orthonormierten Funktionen $\varphi_1, \ldots, \varphi_N$ exakt gelten, und die freie Stützstelle besitzt die Verteilungsdichte $u(\boldsymbol{x}_1)$ in bezug auf das Maß μ. Daher ist die Summe

$\varkappa_n[f]$ genau dann eine erwartungstreue Schätzung des Integrals J, wenn die Abbildungen T_i des Raumes $\mathfrak{X}$ dem Gleichungssystem

$$\sum_{i=1}^{n} \varphi_q(\boldsymbol{x}_i)\, \varphi_1(\boldsymbol{x}_i)\, |\Phi_{i-1}| = 0\,, \qquad q = 2, \ldots, N\,, \tag{4.4.6}$$

(fast überall in bezug auf das μ-Maß) genügen. Die Koeffizienten der Näherungssumme werden dabei durch

$$A_i(Q) = \frac{\varphi_1(\boldsymbol{x}_1)\, |\Phi_{i-1}|}{\sum\limits_{i=1}^{n} \varphi_i^2(\boldsymbol{x}_i)\; |\Phi_{i-1}|}\,, \qquad i = 1, \ldots, n\,, \tag{4.4.7}$$

bestimmt, und $u(\boldsymbol{x}_1)$ hat die Gestalt

$$u(\boldsymbol{x}_1) = \frac{g(\boldsymbol{x}_1)}{n} \sum_{i=1}^{n} \varphi_i^2(\boldsymbol{x}_i)\, |\Phi_{i-1}|\,, \tag{4.4.8}$$

wobei $g(\boldsymbol{x})$ eine beliebige meßbare nichtnegative Funktion mit

$$n^{-1} \sum_{i=1}^{n} g(\boldsymbol{x}_i) = 1 \qquad (\mathrm{mod}\, \mu) \tag{4.4.9}$$

ist.

Beweis. *Notwendigkeit.* Aus der Symmetrie der Summe $\varkappa_n[f]$ und den Beziehungen für T_i ergibt sich

$$A_i(\boldsymbol{x}_1) = A_1(\boldsymbol{x}_i) = A_1(T_{i-1}(\boldsymbol{x}_1)), \qquad i = 1, \ldots, n,$$

und

$$\mathsf{M}\varkappa_n[f] = \int \sum_{i=1}^{n} A_1(T^{i-1}(\boldsymbol{x}_1))\, u(\boldsymbol{x}_1)\, f(T^{i-1}(\boldsymbol{x}_1))\, \mu(d\boldsymbol{x}_1).$$

Nun gilt aber

$$\begin{aligned} &\int A_1(T^{i-1}(\boldsymbol{x}_1))\, f(T^{i-1}(\boldsymbol{x}_1))\, u(\boldsymbol{x}_1)\, \mu(d\boldsymbol{x}_1) \\ &= \int A_1(z)\, f(z)\, u(T^{1-i}(z))\, |\Phi_{1-i}|\, \mu(dz) \\ &= \int A_1(\boldsymbol{x})\, \frac{u(T^{1-i}(\boldsymbol{x}))}{|\Phi_{i-1}|}\, f(\boldsymbol{x})\, \mu(d\boldsymbol{x})\,. \end{aligned}$$

So erhält die Bedingung für die Erwartungstreue von $\varkappa_n[f]$ folgende Gestalt:

$$\mathsf{M}\varkappa_n[f] = \int A_1(\boldsymbol{x})\, f(\boldsymbol{x}) \sum_{=1}^{n} \frac{u(T^{1-i}(\boldsymbol{x}))}{|\Phi_{i-1}|}\, \mu(d\boldsymbol{x}) = \int f(\boldsymbol{x})\, \varphi_1(\boldsymbol{x})\, \mu(d\boldsymbol{x})$$

für eine beliebige Funktion $f \in \mathbf{L}_2(\mu)$. Daraus ergibt sich fast überall

$$A_1(\boldsymbol{x}) \sum_{i=1}^{n} \frac{u(T^{1-i}(\boldsymbol{x}))}{|\Phi_{i-1}|} = \varphi_1(\boldsymbol{x})\,, \qquad \boldsymbol{x} \in \mathfrak{X}\,, \;\; (\mathrm{mod}\; \mu) \tag{4.4.10}$$

Weil aber $T^{1-i} = T^{n-i-1}$ wegen (4.4.4) gilt, ist die letzte Gleichung äquivalent zu

$$A_1(\boldsymbol{x}) = \sum_{m=0}^{n-1} u(T^m(\boldsymbol{x}))\, |\Phi_m| = \varphi_1(\boldsymbol{x})\,. \tag{4.4.11}$$

Aus der Formel für die Funktionaldeterminante einer zusammengesetzten Abbildung ergibt sich

$$|\Phi_i(\boldsymbol{x})| = \prod_{k=1}^{i} |\Phi(T^{k-1}(\boldsymbol{x}))| , \qquad k, i = 1, \ldots, n-1 . \tag{4.4.12}$$

Folglich gilt

$$|\Phi_i(T^k(\boldsymbol{x}))| = \frac{|\Phi_{i+k}(\boldsymbol{x})|}{|\Phi_k(\boldsymbol{x})|}. \tag{4.4.13}$$

Weiterhin gilt

$$\int u(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = \int u(T^i(\boldsymbol{x}))\, |\varphi_i|\, \mu(d\boldsymbol{x}) , \qquad i = 0, \ldots, n-1 ,$$

und daher ist

$$\begin{aligned}\int u(\boldsymbol{x})\, \mu(d\boldsymbol{x}) &= n^{-1} \int \sum_{i=0}^{n-1} u(T^i(\boldsymbol{x}))\, |\Phi_i|\, \mu(d\boldsymbol{x}) \\ &= \int n^{-1} \sum_{m=0}^{n-1} g(T^m(\boldsymbol{x})) \sum_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i)\, |\Phi_{i-1}|\, \mu(d\boldsymbol{x}) = 1 .\end{aligned}$$

Die notwendige Normierungsbedingung für die Funktion $u(\boldsymbol{x})$ ist also erfüllt.

Schließlich ergibt sich

$$\frac{A_1(T^k(\boldsymbol{x}))}{|\Phi_k(\boldsymbol{x})|} \sum_{i=0}^{n-1} u(T^i(\boldsymbol{x}))\, |\Phi_i(\boldsymbol{x})| = \varphi_1(T^k(\boldsymbol{x})), \quad (\text{mod } \mu) \tag{4.4.14}$$

Aus (4.4.10) und (4.4.14) folgt

$$\frac{A_1(\boldsymbol{x})}{A_1(T^k(\boldsymbol{x}))} = \frac{\varphi_1(\boldsymbol{x})}{\varphi_1(T^k(\boldsymbol{x}))\, |\Phi_k(\boldsymbol{x})|}, \qquad k=0, \ldots, n-1 , \quad (\text{mod } \mu) \tag{4.4.15}$$

und damit

$$\varkappa_n[f] = \frac{A_1(\boldsymbol{x}_1)}{\varphi_1(\boldsymbol{x}_1)} \sum_{i=1}^{n} f(\boldsymbol{x}_i)\, \varphi_1(\boldsymbol{x}_i)\, |\Phi_{i-1}| .$$

Soll diese Quadraturformel für die zu $\varphi_1(\boldsymbol{x})$ orthogonalen Funktionen $\varphi_2(\boldsymbol{x}), \ldots, \varphi_N(\boldsymbol{x})$ exakt gelten, so folgt jetzt die Gültigkeit von (4.4.6). Soll die Formel außerdem für $\varphi_1(\boldsymbol{x})$ exakt gelten, so ergibt sich

$$A_1(\boldsymbol{x}_1) = \frac{\varphi_1(\boldsymbol{x}_1)}{\sum_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i)\, |\Phi_{i-1}|}.$$

Setzt man diesen Ausdruck für $A_1(\boldsymbol{x}_1)$ in (4.4.14) ein, so erhält man

$$\sum_{m=0}^{n-1} u(T^m(x))\, |\Phi_m| = \sum_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i)\, |\Phi_{i-1}| . \tag{4.4.16}$$

Wir stellen jetzt $u(\boldsymbol{x}_1)$ in der Form

$$u(\boldsymbol{x}_1) = g(\boldsymbol{x}_1) \sum_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i)\, |\Phi_{i-1}|$$

dar. Dann gilt

$$u(T^m(\boldsymbol{x}_1)) = \frac{g(T^m(\boldsymbol{x}_1))}{|\Phi_m|} \sum_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i) \, |\Phi_{i-1}| \, .$$

Mit (4.4.16) ergibt sich daraus, daß die Funktion $g(\boldsymbol{x})$ der Bedingung

$$\frac{1}{n} \sum_{m=0}^{n-1} g(T^m(\boldsymbol{x})) = 1 \pmod{\mu}$$

fast überall genügt, was mit (4.4.9) gleichwertig ist.

Hinlänglichkeit. Die Abbildung T_i, die $\mathfrak{X}$ in sich abbildet, soll der Forderung (4.4.4) genügen, die Näherungssumme wird durch die Gleichungen (4.4.7) und die Verteilungsdichte der Stützstelle $\boldsymbol{x}_1$ durch die Gleichung (4.4.8) bestimmt. Dann folgt aus (4.4.7) unmittelbar

$$\sum_{i=1}^{n} A_i(Q) \, \varphi_j(\boldsymbol{x}_i) = \begin{cases} 1 & \text{für} \quad j = 1 \, , \\ 0 & \text{für} \quad j = 2, \ldots, N \, . \end{cases}$$

Das bedeutet aber, daß die Quadraturformel für die Funktionen $\varphi_1, \ldots, \varphi_N$ exakt gilt. Schließlich ergibt sich

$$\mathsf{M}\varkappa_n[f] = \int \sum_{i=1}^{n} \varphi_1(\boldsymbol{x}_i) \, |\Phi_{i-1}| \, f(\boldsymbol{x}_i) \, \frac{g(\boldsymbol{x}_1)}{n} \, \mu(d\boldsymbol{x}_1)$$
$$= \int f(\boldsymbol{x}_1) \, \varphi_1(\boldsymbol{x}_1) \, n^{-1} \sum_{i=1}^{n} g(\boldsymbol{x}_i) \, \mu(d\boldsymbol{x}_1) = \int f(\boldsymbol{x}) \, \varphi_1(\boldsymbol{x}) \, \mu(d\boldsymbol{x}) \, .$$

Damit ist der Satz bewiesen.

Für die Herleitung einer stochastischen Quadraturformel mit einer freien Stützstelle, die für gegebene Funktionen φ_i, $i = 1, \ldots, N$, exakt gilt, muß man also solche Lösungen $\boldsymbol{x}_i = T_{i-1}(\boldsymbol{x}_1)$, $i = 2, \ldots, n$, des Gleichungssystems (4.4.2) finden, die den Raum $\mathfrak{X}$ in sich abbilden. Dabei stellen wir fest, daß man als Funktion $g(\boldsymbol{x})$ mit den im Satz genannten Eigenschaften insbesondere

$$g(\boldsymbol{x}) = \frac{\varphi_1^2(\boldsymbol{x})}{\sum\limits_{i=1}^{n} \varphi_1^2(\boldsymbol{x}_i) \, |\Phi_{i-1}|}$$

wählen kann. Dann hat wegen (4.4.8) die dieser Wahl von $g(\boldsymbol{x})$ entsprechende Verteilungsdichte $u(\boldsymbol{x})$ unabhängig von der konkreten Gestalt der Abbildungen T_i die folgende Standardform:

$$u(\boldsymbol{x}) = \varphi_1^2(\boldsymbol{x}) \, .$$

Anschließend werden auf der Grundlage von Satz 4.7 als Beispiel einige Quadraturformeln mit einer freien Stützstelle hergeleitet, die für ein gegebenes Funktionensystem exakt gelten. Wir legen hier immer das Lebesguesche Maß zugrunde, so daß die Forderung nach der Existenz der Funktionaldeterminante der Abbildung T_i in diesem Fall der Forderung nach stückweiser Glattheit dieser Abbildung

äquivalent ist. Daher wird das System (4.4.6) im allgemeinen ein nichtlineares Differentialgleichungssystem sein. Außerdem nehmen wir immer an, daß $\varphi_1(\boldsymbol{x}) = \text{const}$ ist und dementsprechend $u(\boldsymbol{x}) = \varphi_1^2(\boldsymbol{x}) = \text{const}$ gilt.

1. $\mathfrak{X} = [-1, +1]$, $\varphi_1(x) = 1/\sqrt{2}$, $\varphi_2(x) = \sqrt{3/2}\, x$, $n = 2$. Das System (4.4.6) führt hier zu der Differentialgleichung

$$\left|\frac{dx_2}{dx_1}\right| = -\frac{x_2}{x_1}. \tag{4.4.17}$$

Wir suchen stückweise glatte Lösungen der Gleichung (4.4.17). Solche Lösungen sind die drei folgenden

$$x_2 = \begin{cases} (c^2 - x_1^2)^{1/2} & \text{für} \quad x_1 \in [-c, 0)\,, \\ -(c^2 - x_1^2)^{1/2} & \text{für} \quad x_1 \in [0, c]\,, \end{cases} \tag{4.4.18}$$

$$x_2 = \begin{cases} (c^2 + x_1^2)^{1/2} & \text{für} \quad x_1 \leqq 0\,, \\ -(c^2 + x_1^2)^{1/2} & \text{für} \quad x_1 > 0\,, \end{cases} \tag{4.4.19}$$

$$x_2 = \begin{cases} (x_1^2 - c^2)^{1/2} & \text{für} \quad x_1 \leqq -c\,, \\ -(x_1^2 - c^2)^{1/2} & \text{für} \quad x_1 > c \end{cases} \tag{4.4.20}$$

mit einer beliebigen Konstanten $c \geqq 0$.

Weitere Lösungen des Systems (4.4.17) sind verschiedene stückweise stetige Kurven, die aus einzelnen Teilkurven dieser drei Lösungen zusammengesetzt sind, so daß die sich ergebende Funktion umkehrbar eindeutig ist. Ein Beispiel ist etwa

$$x_2 = \begin{cases} (c^2 - x_1^2)^{1/2} & \text{für} \quad x_1 \in [-c, 0)\,, \\ -(c^2 + x_1^2)^{1/2} & \text{für} \quad x_1 > 0\,. \end{cases}$$

Wählt man $c = 1$ für (4.4.18) und $c = 0$ für (4.4.19) und (4.4.20), so erhält man die folgenden beiden Lösungen von (4.4.17):

$$x_2 = \begin{cases} (1 - x_1^2)^{1/2} & \text{für} \quad x_1 \in [-1, 0)\,, \\ -(1 - x_1^2)^{1/2} & \text{für} \quad x_1 \in [0, 1] \end{cases}$$

und

$$x_2 = -x_1 \qquad \text{für} \qquad x_1 \in [-1, +1]\,,$$

die beide eine umkehrbare eindeutige Abbildung des Intervalls $[-1, +1]$ auf sich definieren und der Symmetriebedingung (4.4.4) genügen.

Die entsprechenden Näherungssummen haben die Gestalt

$$\varkappa_2^{(1)}[f] = \frac{2}{|x_1| + |x_2|}\,[|x_2|\, f(x_1) + |x_1|\, f(x_2)] \tag{4.4.21}$$

mit $x_1^2 + x_2^2 = 1$, $x_1 x_2 < 0$ und

$$\varkappa_2^{(2)}[f] = f(x_1) + f(-x_1)\,. \tag{4.4.22}$$

Die der Näherungssumme (4.4.22) entsprechende Quadraturformel ist identisch mit der Formel (4.3.1). Wie schon in § 3 bemerkt wurde, besteht in der Anwendung dieser Formel und ihrer Verallgemeinerungen (Hammersley und Morton [1], Morton [1]) der Inhalt der Methode der symmetrischen Stichprobe (*antithetic*

variate). Die Monographie von HAMMERSLEY und HANDSCOMB [1] enthält eine ausführliche Übersicht über Arbeiten, die zu dieser Methode Beziehung haben, sowie die entsprechende Bibliographie.

Die der Summe (4.4.21) entsprechende Formel erhielt als erster GRANOWSKI [2]. Wie schon bemerkt, gelten beide Formeln exakt für $f \equiv 1$ und $f \equiv x$. Vergleicht man die Summen (4.4.21) und (4.4.22) mit (4.2.20), so stellt man leicht fest, daß sie (besonders (4.4.22)) wesentlich einfacher anzuwenden sind als (4.2.20).

2. $\mathfrak{X} = [-1, +1]$, $\varphi_i(x)$, $i = 1, 2, 3, 4$, sind die normierten Legendreschen Polynome vom Grad $i - 1$. Weil φ_2 und φ_4 ungerade Funktionen sind, erfüllen wir mit $x_1 = -x_2 = x$ und $x_3 = -x_4 = y$ die erste und dritte Gleichung des Systems (4.4.6) bei beliebiger Abhängigkeit zwischen x und y. Das System (4.4.6) wird so auf eine Differentialgleichung

$$\left|\frac{dy}{dx}\right| = -\frac{\varphi_3(x)}{\varphi_3(y)} \equiv -\frac{3x^2 - 1}{3y^2 - 1}$$

zurückgeführt. Eine ihrer Lösungen lautet

$$y = \begin{cases} (\frac{1}{2})\left(x + \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [-1, 0), \\ (\frac{1}{2})\left(x - \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [0, 1]. \end{cases} \tag{4.4.23}$$

Die entsprechende Näherungssumme hat die Gestalt

$$\varkappa_4[f] = \frac{1}{3\left(x + \sqrt{4 - 3x^2}\right)} \{[f(x) + f(-x)]\, 2\sqrt{4 - 3x^2} + [f(y) + f(-y)]\,[3x + \sqrt{4 - 3x^2}]\} \tag{4.4.24}$$

wobei y durch (4.4.23) definiert wird. Die Abbildung (4.4.23) genügt allerdings nicht den Symmetriebedingungen (4.4.4). Um die Erfüllung dieser Bedingungen zu sichern, kann man eine entsprechende Abbildung T in folgender Weise definieren:

$$T(x) = \begin{cases} (\frac{1}{2})\left(x + \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [-1, -\frac{1}{\sqrt{3}}), \\ -(\frac{1}{2})\left(x + \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [-\frac{1}{\sqrt{3}}, 0), \\ -(\frac{1}{2})\left(x - \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [0, \frac{1}{\sqrt{3}}), \\ (\frac{1}{2})\left(x - \sqrt{4 - 3x^2}\right) & \text{für} \quad x \in [\frac{1}{\sqrt{3}}, 1]. \end{cases}$$

Dann ist $T^2 = -T^0$ und folglich $T^4 = T^0$. Die symmetrische Näherungssumme, die auf der Grundlage dieser Abbildungen aufgestellt wird, fällt allerdings (bis auf die Numerierung der Stützstellen) mit der Näherungssumme (4.4.24) zusammen. Beide Quadraturformeln sind also äquivalent. Wir bemerken noch, daß sich aus (4.4.24) für $x = 1$ die bekannte Simpsonsche Formel ergibt.

3. Es sei $\mathfrak{X}$ ein Gebiet des s-dimensionalen euklidischen Raumes, das symmetrisch in bezug auf den Koordinatenursprung gelegen ist. Wenn $|\mathfrak{X}|$ das Volumen des Gebietes $\mathfrak{X}$ bezeichnet, dann gilt die Quadraturformel

$$\int_{\mathfrak{X}} f(\boldsymbol{x})\, d\boldsymbol{x} \approx \frac{|\mathfrak{X}|}{2} [f(\boldsymbol{x}_1) + f(-\boldsymbol{x}_1)] \tag{4.4.25}$$

exakt für Konstanten und für alle Polynome ersten Grades in s Veränderlichen. Die Stützstelle $\boldsymbol{x}_1$ ist im Gebiet $\mathfrak{X}$ gleichmäßig verteilt. Diese Formel stellt das s-dimensionale Analogon der Formel (4.3.1) dar (vgl. MORTON [1]).

4. Die Menge $\mathfrak{X}$ sei der Einheitskreis $x_1^2 + x_2^2 = 1$. Als Funktionen φ_i, $i = 1, 2, 3, 4$, werden $c_1, c_2x_1, c_3x_2, c_4x_1x_2$ gewählt, wobei die c_i geeignete Normierungskonstanten sind.

Wir suchen nun die einfachste symmetrische stochastische Quadraturformel mit einer freien Stützstelle

$$\int f(\boldsymbol{x})\, d\boldsymbol{x} \approx \frac{|\mathfrak{X}|}{4} \sum_{i=1}^{4} f(T^{i-1}(\boldsymbol{x}))\,. \tag{4.4.26}$$

GRANOWSKI [1] zeigt in einer Arbeit, daß die gestellte Aufgabe durch eine orthogonale Abbildung T gelöst wird, deren Matrix folgende Gestalt hat (wir bezeichnen sie mit demselben Buchstaben T):

$$\begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

T stellt also eine Drehung um den Winkel 120° dar. Dann ist

$$\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

und daher $|\Phi_i| \equiv 1$, $i = 1, 2$. Man kann zeigen, daß die dieser Abbildung T entsprechende Quadraturformel die einzige Lösung der gegebenen Aufgabe darstellt, die die Gestalt (4.4.25) hat. In derselben Arbeit wird dieses Resultat auch auf den Fall einer s-dimensionalen Kugel verallgemeinert.

Wir wollen nun ausführlicher solche stochastischen Quadraturformeln mit einer freien Stützstelle studieren, die für $\varphi_0(\boldsymbol{x}) \equiv 1$ exakt gelten. Nach (4.4.7) und (4.4.8) haben solche Formeln folgende Gestalt:

$$\int f(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \approx \frac{\mu(\mathfrak{X})}{\sum\limits_{i=0}^{n-1} |\Phi_i|} \sum_{i=0}^{n-1} f(T^i(\boldsymbol{x}))\, |\Phi_i|\,, \tag{4.4.27}$$

wobei die Stützstelle $\boldsymbol{x}$ in $\mathfrak{X}$ mit der Dichte

$$u(\boldsymbol{x}) = \frac{g(\boldsymbol{x})}{n\mu(\mathfrak{X})} \sum_{i=0}^{n-1} |\Phi_i| \tag{4.4.28}$$

in bezug auf das Maß μ verteilt ist. Die Abbildung T genügt dabei den Bedingungen (4.4.4) und die Funktion $g(\boldsymbol{x}) \gtreqless 0$, $x \in \mathfrak{X}$, der Bedingung (4.4.9).

Um eine Abbildung T mit den gewünschten Eigenschaften zu finden, wird die folgende Konstruktion angewendet. Es werden Teilmengen $d_i \subset \mathfrak{A}$, $i = 1, \ldots, n$, mit folgenden Eigenschaften betrachtet:

$$\mu(d_i) > 0\,, \qquad \mu(d_i \cap d_j) = 0\,, \qquad \mu\left(\bigcup_{i=1}^{n} d_i\right) = \mu(\mathfrak{X})\,, \tag{4.4.29}$$
$$i \neq j; \quad i, j = 1, \ldots, n\,.$$

Nun sei T eine solche Abbildung der Menge $\mathfrak{X}$, die jede Menge d_{i_l} auf eine Menge d_{i_k} $(k \neq l)$ in Übereinstimmung mit einer gewissen zyklischen Permutation

$$\begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}$$

der Länge n abbildet. Dann bilden die Abbildungen T^i, $i = 0, \dots, n-1$, eine zyklische Gruppe $\mathcal{M}$ der Ordnung n. Dabei gelten für die Funktionaldeterminanten $|\Phi_i|$ die Formeln (4.4.12) und (4.4.13). Aus diesen Formeln folgt, daß die der Beziehung (4.4.27) entsprechenden Näherungssummen $\varkappa_n[f]$ invariant bezüglich der Abbildung T sind:

$$\varkappa_n[f; \boldsymbol{x}_n] = \varkappa_n[f; T(\boldsymbol{x})] . \tag{4.4.30}$$

Für jede meßbare Funktion g werden wir die Größe

$$\frac{1}{n} \sum_{i=0}^{n-1} g\big(T^i(\boldsymbol{x})\big) = \langle g(\boldsymbol{x}) \rangle_{\mathcal{M}}$$

Mittelwert der Funktion g im Punkt $\boldsymbol{x}$ nach der Gruppe $\mathcal{M}$ nennen.

Die Bedingung (4.4.9) kann daher in der Form

$$\langle g(\boldsymbol{x}) \rangle_{\mathcal{M}} = 1 \qquad \text{(fast überall)} \tag{4.4.31}$$

geschrieben werden. Aus (4.4.30) und (4.4.31) folgt, daß die Momente $\mathsf{M}[\varkappa_n[f, \boldsymbol{x}_{u_g}]]^q$, $q = 1, 2, \dots$, für alle Dichten u_g der Gestalt (4.4.28) gleich sind. Insbesondere kann man

$$g(\boldsymbol{x}) = n \left(\sum_{i=0}^{n-1} |\Phi_i(\boldsymbol{x})| \right)^{-1}$$

setzen. Hier folgt die Richtigkeit der Bedingung (4.4.31) unmittelbar aus (4.4.12).

Bei einer solchen Wahl der Funktion g erhalten wir $u(x) = \big(\mu(\mathfrak{X})\big)^{-1}$, so daß die freie Stützstelle $\boldsymbol{x}$ immer als gleichverteilt in $\mathfrak{X}$ angenommen werden kann. Das wird auch weiterhin vorausgesetzt. Wir betrachten nun eine auf $\mathfrak{X}$ gegebene Funktion φ der Gestalt

$$\varphi(\boldsymbol{x}) = h(\boldsymbol{x}) \sum_{i=1}^{n} |\Phi_i| , \tag{4.4.32}$$

wobei h eine beliebige Funktion mit $\langle h \rangle_{\mathcal{M}} = \text{const}$ ist. Dann gilt auf Grund von (4.4.12)

$$\varphi\big(T^q(\boldsymbol{x})\big) = \frac{h\big(T^q(\boldsymbol{x})\big)}{|\Phi_q|} \sum_{i=0}^{n-1} |\Phi_i|$$

und fast überall

$$\varkappa_n[\varphi; \boldsymbol{x}] = \mu(\mathfrak{X}) \langle h \rangle_{\mathcal{M}} = \text{const} . \tag{4.4.33}$$

Die Gleichung (4.4.33) besagt, daß die Quadraturformel (4.4.27) für alle Funktionen φ der Form (4.4.32) aus $\mathbf{L}_1(\mu)$ exakt gilt. Man beweist auch leicht die umgekehrte Behauptung: Jede Funktion aus $\mathbf{L}_1(\mu)$, für die die Quadraturformel (4.4.27) exakt gilt, ist in der Gestalt (4.4.32) darstellbar. Daher gilt der

Satz 4.8. *Eine symmetrische Quadraturformel* (4.4.27) *mit einer freien Stützstelle gilt für genau die Funktionen aus* $\mathbf{L}_1(\mu)$ *exakt, die in der Form* (4.4.32) *darstellbar sind.*

Satz 4.8 stellt eine unmittelbare Verbindung her zwischen der gegebenen zyklischen Gruppe von Abbildungen T^i und der Funktionenmenge, für die die Quadraturformel (4.4.27) exakt gilt. Das Resultat dieses Satzes liegt dem Beweis des nächsten Satzes 4.9 zugrunde.

Eine wichtige Eigenschaft der stochastischen Quadraturformeln mit einer freien Stützstelle ist die Zulässigkeit der ihnen entsprechenden Prozeduren in einer genügend umfangreichen Klasse von (Integrations-)Prozeduren. Zur Vereinfachung der Darstellung werden wir jetzt auch die Sprechweise „Zulässigkeit einer Quadraturformel" gebrauchen. Dann gilt der

Satz 4.9. *Die stochastische Quadraturformel* (4.4.27) *mit einer freien Stützstelle ist in der Klasse aller stochastischen Quadraturformeln für die Klasse* $\mathbf{L}_2(\mu)$ *zulässig.*

Beweis. Wir nehmen das Gegenteil an. Dann gibt es entsprechend der Definition der Zulässigkeit eine Prozedur $(\tilde{\varkappa}_n[f], \tilde{u})$, die über die gegebene Prozedur $(\varkappa_n[f], u)$ auf der ganzen Klasse $\mathbf{L}_2(\mu)$ dominiert und die wie die betrachtete Prozedur folglich auch für dieselben Funktionen $f \in \mathbf{L}_2(\mu)$ exakt gilt.

a) Wir zeigen zunächst, daß in diesem Fall $\tilde{\varkappa}_n[f] \approx J$ für alle die Funktionen φ der Gestalt (4.4.32) mit

$$|h(\boldsymbol{x})| \leqq C < \infty\,, \qquad \boldsymbol{x} \in \mathfrak{X} \qquad \text{(fast überall)} \tag{4.4.34}$$

exakt gelten muß. (Es ist zu bemerken, daß solche Funktionen φ zu $\mathbf{L}_1(\mu)$, aber nicht unbedingt zu $\mathbf{L}_2(\mu)$ gehören.)

Wir betrachten nun eine Folge von Funktionen $\{\varphi_N\}$:

$$\varphi_N(\boldsymbol{x}) = \begin{cases} \varphi(\boldsymbol{x}) & \text{für} \quad |\varphi(\boldsymbol{x})| \leqq N\,, \\ N & \text{für} \quad \varphi(\boldsymbol{x}) > N\,, \\ -N & \text{für} \quad \varphi(\boldsymbol{x}) < -N\,, \end{cases} \tag{4.4.35}$$

wobei φ eine beliebige Funktion der Gestalt (4.4.32) ist. Aus (4.4.34) und (4.4.35) folgt

$$\varphi_N(\boldsymbol{x}) \to \varphi(\boldsymbol{x}) \qquad \text{fast überall} \quad \text{für } N \to \infty\,, \tag{4.4.36}$$

$$\varphi_N \leqq |\varphi| \in \mathbf{L}_1(\mu)\,, \qquad N = 1, 2, \ldots \tag{4.4.37}$$

Daraus ergibt sich nach dem Lebesgueschen Satz

$$\int \varphi_N \mu(d\boldsymbol{x}) \to \int \varphi \mu(d\boldsymbol{x}) \qquad \text{für} \qquad N \to \infty\,. \tag{4.4.38}$$

Weil dabei $|\varphi_N| \leqq N$ ist, gilt $\varphi_N \in \mathbf{L}_2(\mu)$. Daher müssen in Übereinstimmung mit den von uns gemachten Voraussetzungen über $(\tilde{\varkappa}_n[f], \tilde{u})$ folgende Aussagen gelten:

$$\mathsf{M}[\tilde{\varkappa}_n[\varphi_N; \boldsymbol{x}_{\tilde{u}}]] = \int \varphi_N(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{4.4.39}$$

und

$$\mathsf{D}[\tilde{\varkappa}_n[\varphi_N; \boldsymbol{x}_{\tilde{u}}]] \leqq \mathsf{D}[\varkappa_n[\varphi_N; \boldsymbol{x}_u]]\,, \qquad N = 1, 2, \ldots \tag{4.4.40}$$

Weiterhin folgt aus (4.4.27), (4.4.37) und (4.4.32) bis (4.4.34)

$$\varkappa_n[\varphi_N] \leqq \varkappa_n[|\varphi|] = n\mu(\mathfrak{X}) \langle |h| \rangle_{\mathcal{M}} \leqq C\mu(\mathfrak{X}) , \tag{4.4.41}$$

$$\varkappa_n[\varphi_N] \to \varkappa_n[\varphi] \quad \text{für} \quad N \to \infty , \tag{4.4.42}$$

$$\tilde{\varkappa}_n[\varphi_N] \to \tilde{\varkappa}_n[\varphi] \quad \text{für} \quad N \to \infty . \tag{4.4.43}$$

So ergibt sich nach (4.4.41) und (4.4.42) und unter Benutzung des Lebesgueschen Satzes

$$\mathsf{M}[\varkappa_n^2[\varphi_N ; \boldsymbol{x}_u]] \to \mathsf{M}[\varkappa_n^2[\varphi ; \boldsymbol{x}_u]] .$$

Da die Quadraturformel $\varkappa_n[f] \approx J$ exakt für Funktionen φ gilt, folgt hieraus und aus (4.4.38) für $N \to \infty$

$$\mathsf{D}[\varkappa_n[\varphi_N ; \boldsymbol{x}_u]] \to 0 .$$

Nach (4.4.40) ergibt das

$$\mathsf{D}[\tilde{\varkappa}_n[\varphi_N ; \boldsymbol{x}_{\tilde{u}}]] \to 0$$

für $N \to \infty$, woraus in Übereinstimmung mit (4.4.39) die Existenz einer Funktionenfolge $\{\varphi_{n_q}\}$ mit

$$\tilde{\varkappa}_n[\varphi_{n_q}] - \int \varphi_{n_q} \mu(d\boldsymbol{x}) \to 0$$

für $n_q \to \infty$ folgt. Zusammen mit (4.4.43) und (4.4.38) kann man daraus schließen, daß $\tilde{\varkappa}_n[\varphi] = \int \varphi \mu(d\boldsymbol{x})$ ist. Damit ist die Behauptung des Teiles a) bewiesen.

b) Wir setzen in (4.4.32) $h(\boldsymbol{x}) = h_i(\boldsymbol{x})$ mit

$$h_i(\boldsymbol{x}) = \begin{cases} 0 & \text{für} \quad \boldsymbol{x} \notin d_i , \\ 1 & \text{für} \quad \boldsymbol{x} \in d_i , \end{cases} \qquad i = 1, \ldots, n . \tag{4.4.44}$$

Weil $\langle h_i \rangle_{\mathcal{M}} = 1/n$ ist und die Funktionen h_i der Bedingung (4.4.34) genügen, muß gemäß a) die Formel $\tilde{\varkappa}_n[f] \approx J$ exakt für die diesen h_i entsprechenden Funktionen φ_i der Gestalt (4.4.32) gelten. Daraus folgt, daß das Verteilungsgesetz $\tilde{U}$ der Stützstellen in der stochastischen Näherungssumme $\tilde{\varkappa}_n$ so beschaffen sein muß, daß in jede Menge d_i, $i = 1, \ldots, n$, mit der Wahrscheinlichkeit 1 eine der n stochastischen Stützstellen x_i fällt, da andernfalls für ein gewisses i mit positiver Wahrscheinlichkeit

$$\tilde{\varkappa}_n[\varphi_i] = 0 \neq \int \varphi_i \mu_i(d\boldsymbol{x})$$

gelten würde. Da die Quadraturformel für die Funktionen φ_i exakt gelten soll, erhalten wir daher aus (4.4.32) und (4.4.3)

$$\tilde{A}_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \left(\sum_{i=0}^{n-1} |\Phi_i| \right)_{\boldsymbol{x} = \boldsymbol{x}_q} = \mu(\mathfrak{X}) .$$

Daraus folgt, daß die Näherungssumme $\tilde{\varkappa}_n[f]$ folgende Gestalt haben muß:

$$\tilde{\varkappa}_n[f] = \mu(\mathfrak{X}) \sum_{q=1}^{n} \frac{f(\boldsymbol{x}_q)}{\left(\sum\limits_{i=0}^{n-1} |\Phi_i| \right)_{\boldsymbol{x} = \boldsymbol{x}_q}} , \tag{4.4.45}$$

wobei man ohne Beschränkung der Allgemeinheit annehmen kann, daß

$$\boldsymbol{x}_q \in d_q = T^{q-1}(d_1)\,, \qquad q = 1, \ldots, n\,, \tag{4.4.46}$$

gilt. Ist daher φ eine beliebige der in a) betrachteten Funktionen dann muß nach (4.4.32)

$$\tilde{\varkappa}_n[\varphi] = \mu(\mathfrak{X}) \sum_{q=1}^{n} h(\boldsymbol{x}_q) = \text{const} = n\mu(\mathfrak{X})\, \langle h \rangle_{\mathcal{M}} \tag{4.4.47}$$

gelten. Es sei jetzt

$$h_{1,q}(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x}) & \text{für} \quad \boldsymbol{x} \in d_1\,, \\ -h\big(T^{1-q}(\boldsymbol{x})\big) & \text{für} \quad \boldsymbol{x} \in d_q\,, \\ 0 & \text{für} \quad \boldsymbol{x} \notin d_1 \cup d_q \end{cases}$$

mit $|h(\boldsymbol{x})| \leqq c < \infty$, $\boldsymbol{x} \in \mathfrak{X}$, Dann ist $\langle h \rangle_{\mathcal{M}} = 0$, und aus (4.4.47) und (4.4.46) folgt die Beziehung (fast überall bezüglich $\tilde{U}$)

$$h_{1,q}(\boldsymbol{x}_1) + h_{1,q}(\boldsymbol{x}_q) = h(\boldsymbol{x}_1) - h\big(T^{1-q}(\boldsymbol{x}_1)\big) = 0$$

für alle beschränkten Funktionen h. Aber das ist nur dann möglich, wenn $\boldsymbol{x}_q = T^{q-1}(\boldsymbol{x}_1)$, $q = 2, \ldots, n$, ist. Daraus folgt dann wegen (4.4.45) und (4.4.31), daß $\tilde{\varkappa}_n[f]$ und $\varkappa_n[f]$ zusammenfallen müssen. Weil $\tilde{u}$ in der Form

$$\tilde{u}(\boldsymbol{x}) = \frac{\tilde{g}(\boldsymbol{x})}{n\mu(\mathfrak{X})} \sum_{i=1}^{n} |\Phi_i|$$

mit meßbarer Funktion $\tilde{g}(\boldsymbol{x})$, $\tilde{g}(\boldsymbol{x}) > 0$, dargestellt werden kann, folgt aus der Erwartungstreue genau wie im Beweis des Satzes 4.7, daß $\tilde{g}$ die Bedingung $\langle \tilde{g} \rangle_{\mathcal{M}} = 1$ erfüllen muß. Damit ist Teil b) und zugleich der ganze Satz bewiesen.

Wir nehmen nun an, daß $h(\boldsymbol{x})$ eine beliebige Funktion aus $\mathbf{L}_2(\mu)$ ist. Wir setzen $h_1(\boldsymbol{x}) = h(\boldsymbol{x}) - \langle h(\boldsymbol{x}) \rangle_{\mathcal{M}}$ und $h_2(\boldsymbol{x}) = \langle h \rangle_{\mathcal{M}}$. Dann ist $h(\boldsymbol{x})$ in der Form $h(\boldsymbol{x}) = h_1(\boldsymbol{x}) + h_2(\boldsymbol{x})$ mit $\langle h_1 \rangle_{\mathcal{M}} = 0$ darstellbar, wobei $h_2(\boldsymbol{x})$ invariant bezüglich der Abbildung T ist, also $h_2\big(T(\boldsymbol{x})\big) = h_2(\boldsymbol{x})$ fast überall gilt. Aus $|\Phi_0| = 1$ folgt dann

$$\sum_{i=0}^{n-1} |\Phi_i| \geqq 1\,.$$

Daher ist eine beliebige Funktion $f(\boldsymbol{x}) \in \mathbf{L}_2(\mu)$ in der Form

$$f(\boldsymbol{x}) = \frac{1}{n}\big(h_1(\boldsymbol{x}) + h_2(\boldsymbol{x})\big) \sum_{i=1}^{n} |\Phi_i|$$

mit $h_1 + h_2 \in \mathbf{L}_2(\mu)$ darstellbar.

Mit Satz 4.8 und der Invarianz der Funktion h_2 ergibt sich daraus

$$\varkappa_n[f; \boldsymbol{x}] = \mu(\mathfrak{X})\, h_2(\boldsymbol{x})\,.$$

Aus dieser Beziehung folgt unmittelbar

Satz 4.10. *Die Dispersion einer Näherungssumme der Gestalt* (4.4.27) *mit einer freien Stützstelle ist gleich dem Produkt aus dem Maß der Menge* $\mathfrak{X}$ *mit der in der Metrik von* $\mathbf{L}_2(\mu)$ *gemessenen Entfernung von der Funktion* $h_2(\boldsymbol{x})$ *(die in die Dar-*

stellung (4.4.32) *des Integranden f eingeht*) *bis zum Unterraum jener Funktionen aus* $\mathbf{L}_2(\mu)$, *für die die entsprechende Formel exakt gilt.*

Beweis. Es ist

$$\mathsf{D}\varkappa_n[f; \boldsymbol{x}] = \mathsf{M}(h_2(\boldsymbol{x})\,\mu(\mathfrak{X}))^2 - \mu^2(\mathfrak{X})\,\mathsf{M}^2(h_2(\boldsymbol{x}))$$
$$= \mu(\mathfrak{X})\left\{\int h_2^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) - \frac{1}{\mu(\mathfrak{X})}\left[\int h_2(\boldsymbol{x})\,\mu(d\boldsymbol{x})\right]^2\right\}.$$

Außerdem gilt $(h_2, \varphi) = 0$, wobei $\varphi \neq$ const eine beliebige Funktion aus $\mathbf{L}_2(\mu)$ ist, für die die Quadraturformel exakt gilt. Das folgt unmittelbar aus der Definition der Funktionen h_2 und φ.

Damit ist der Satz bewiesen.

Folgerung. Ist $|\Phi_0| = |\Phi_1| = \cdots = |\Phi_{n-1}|$, so ist die Dispersion der Näherungssumme mit einer freien Stützstelle gleich

$$\mu(\mathfrak{X})\left\{\int\left[f(\boldsymbol{x}) - \sum_i a_i\psi_i(\boldsymbol{x})\right]^2 \mu(d\boldsymbol{x})\right\}$$

mit

$$a_i = \int f(\boldsymbol{x})\,\psi_i(\boldsymbol{x})\,\mu(d\boldsymbol{x})\,,$$

wobei die ψ_i orthonormierte Funktionen sind, für die die entsprechende Formel exakt gilt.

Wir kommen nun noch einmal auf die Näherungssumme (4.4.22) zurück und finden dann, daß ihre Dispersion gleich dem Quadrat des Abstandes (in der $\mathbf{L}_2([-1, +1])$-Metrik gemessen) der Funktion f zum Unterraum aller ungeraden quadratisch integrierbaren Funktionen ist. Was die Summe (4.4.21) betrifft, so ist hier die Beschreibung aller Funktionen, für die die entsprechende Formel exakt gilt, eine kompliziertere Aufgabe. Beispielsweise gilt sie auch exakt für die gerade Funktion $f(x) = x^2 - |x|\sqrt{1 - x^2}$.

Ziehen wir abschließend eine Bilanz der erhaltenen Resultate, so ist folgendes zu bemerken. In Abschnitt 1 dieses Paragraphen wurde die Beziehung zwischen der Eigenschaft einer Quadraturformel mit einer freien Stützstelle, exakt für ein gewisses Funktionensystem $\{\varphi_i\}$ zu sein, und den Abbildungen T_i festgestellt. Dabei geben die Abbildungen T_i die Beziehungen zwischen den Stützstellen der Formel wieder. Die Resultate des Abschnitts 2 ermöglichen die Herleitung der umgekehrten Beziehung zwischen der Abbildung T und der Eigenschaft der Quadraturformel, exakt für gewisse Funktionen zu sein. Praktisch ist sowohl die Aufgabe, die Abbildung T aus gegebenen Funktionen φ_i zu bestimmen, als auch die Aufgabe, alle Funktionen tatsächlich anzugeben, für die die Quadraturformel bei gegebenem T exakt gilt, von recht komplizierter Natur. In realen Fällen ist es am einfachsten, diese Verfahren zu kombinieren. Man wählt eine solche Menge von Gleichungen (4.4.2), daß dieses System genügend viele Lösungen hat und sucht darunter eine, wobei gewisse Invarianzeigenschaften ausgenutzt werden. Es ist der Fall möglich, daß die exakte Gültigkeit der Quadraturformel für die von uns gewählten Funktionen mit einer Abbildung T' erreicht wird, die eine zyklische Gruppe der Ordnung m $(m < n)$ bildet. Dann verschwinden $n - m$ Funktionaldeterminanten identisch. In diesem Fall muß die Quadraturformel m Summanden enthalten.

Von gewissem Interesse ist auch die Gegenüberstellung von Methoden zur Einführung des Integrals über abstrakten Maßen (Kap. I), von Quadraturformeln, die exakt für Haarsche Funktionen gelten, und von Quadraturformeln mit einer freien Stützstelle. In allen Fällen spielt bei der Betrachtung die Zerlegung der Ausgangsmenge $\mathfrak{X}$ in durchschnittsfremde Teilmengen eine Rolle. Offensichtlich ist auch die Beziehung zwischen den Quadraturformeln mit einer freien Stützstelle und der Ergodentheorie.

Weitere Beziehungen und Analogien sind leicht aufzuzeigen. Bekannt ist die Rolle, die in der mathematischen Statistik die bezüglich einer Gruppe invarianten Abbildungen spielen. Andererseits wurde die Eigenschaft der Invarianz für die Herleitung der klassischen Quadraturformeln benutzt (Sobolew [1]). Schließlich ist klar, daß die Integrationsverfahren mit Hilfe einer stochastischen Quadraturformel mit einer freien Stützstelle als direkte Verallgemeinerung der Methode der symmetrischen Stichprobe zu betrachten sind.

§ 5. Andere Quadraturformeln, sukzessive Verfahren

Der Standpunkt, den wir bei der Untersuchung der stochastischen Quadraturformeln einnehmen, unterscheidet sich vom klassischen in der Hinsicht, daß das grundlegende Ziel bei der Anwendung einer stochastischen Quadraturformel in der Dispersionsverkleinerung besteht. Die Funktionenklasse $L_2(\mu)$, für die die Formeln Anwendung finden, ist sehr umfangreich. Wie wir sehen, kann man in diesem Fall nicht von der Optimalität irgendeiner Quadraturformel sprechen, sondern die natürliche Eigenschaft einer Formel ist ihre Zulässigkeit. Wenn über die zu integrierende Funktion vor Beginn der Rechnung nichts bekannt ist, dann gibt es im allgemeinen keine andere Empfehlung als die, die zulässige Prozedur zu benutzen. Unter den zulässigen Prozeduren wird man dabei offensichtlich diejenige auswählen, die sich am einfachsten auf einer EDVA realisieren läßt. Ist aber der Aufwand für die Funktionswerteberechnung sehr klein, so kann man auf die Forderung nach der Zulässigkeit verzichten und nur die Forderung nach der einfachen Realisierbarkeit beibehalten, was dann zur gewöhnlichen Monte-Carlo-Methode führt.

Worin besteht nun aber die eigentliche Bedeutung der Behauptung „über die zu integrierende Funktion ist nichts bekannt"? Sie hat einen bestimmten Sinn, wenn die Funktionswerte als Resultat eines Experiments erhalten werden. Wird aber bereits ein Programm zur Funktionswerteberechnung von $f(x)$ aufgestellt, dann gibt es immer zusätzliche Informationen. Den Fall, daß es Informationen über die Glattheit der Funktion $f(x)$ gibt, betrachten wir nicht. Diesen Fragen ist eine umfangreiche Literatur gewidmet, und wir sind in § 1 darauf eingegangen. Eine andere Art von Informationen führt gewöhnlich zu der Aussage, daß die zu integrierende Funktion einer oder einigen genügend einfach zu berechnenden Funktionen ähnlich ist. Diese Information kann recht gut durch eine entsprechende Interpolationsquadraturformel berücksichtigt werden. Dabei wird gewöhnlich eine Formel mit einer kleinen Anzahl von Stützstellen hergeleitet, aber die Funktionen, für die die Formel exakt gilt, können sehr verschiedenartig sein. Ein solcher Standpunkt wurde auch im vorangehenden Paragraphen eingenommen.

Wir waren bestrebt, eine Anzahl genügend einfacher Quadraturformeln zu erhalten, die auf Funktionen aus $\mathbf{L}_2(\mu)$ anwendbar sind. Wie schon bemerkt, ist ein solches Vorgehen eng mit Aufgaben der Versuchsplanung verknüpft. Ausführlicher kommen wir darauf auch in Kapitel V zurück. Anschließend gehen wir auf einige Verfahren der näherungsweisen Integration ein, die bisher in diesem Kapitel noch nicht betrachtet wurden.

1. Quadraturformeln mit teilweise festen Stützstellen. Es sei $f(\boldsymbol{x}) \in \mathbf{L}_2(\mu)$, $\varphi_1(\boldsymbol{x}) = \text{const}$ und $\varphi_2(\boldsymbol{x})$ orthogonal zu $\varphi_1(\boldsymbol{x})$ und normiert. Die Quadraturformel

$$J \equiv \int_{\mathfrak{X}} f(\boldsymbol{x})\,\mu(d\boldsymbol{x}) \approx \frac{\mu(\mathfrak{X})}{\varphi_2(\boldsymbol{x}_2) - \varphi_2(\boldsymbol{x}_1)}\,[\varphi_2(\boldsymbol{x}_2)\,f(\boldsymbol{x}_1) - \varphi_2(\boldsymbol{x}_1)\,f(\boldsymbol{x}_2)] \qquad (4.5.1)$$

gilt exakt für $\varphi_1(\boldsymbol{x})$ und $\varphi_2(\boldsymbol{x})$. Die Menge $\{\boldsymbol{x} : \varphi_2(\boldsymbol{x}) > 0\}$ hat offensichtlich ein von Null verschiedenes Maß. Wir nehmen an, daß $\boldsymbol{x}^* \in \mathfrak{X}$ der Punkt ist, in dem die Funktion $\varphi_2(\boldsymbol{x})$ ihr Supremum erreicht, wobei $\mu(\boldsymbol{x} : \varphi_2(\boldsymbol{x}) = \varphi_2(\boldsymbol{x}^*)) = 0$ gelten soll und $f(\boldsymbol{x})$ in diesem Punkt definiert ist. Wir setzen $\boldsymbol{x}_2 = \boldsymbol{x}^*$. Man sieht leicht, daß die der Beziehung (4.5.1) entsprechende Näherungssumme $\varkappa_2[f]$ eine erwartungstreue Schätzung des Integrals J liefert, wenn $\boldsymbol{x}_2 = \boldsymbol{x}^*$ ist und $\boldsymbol{x}_1$ mit der Dichte

$$u(\boldsymbol{x}_1) = \frac{1}{\varphi_2(\boldsymbol{x}^*)\,\mu(\mathfrak{X})}\,[\varphi_2(\boldsymbol{x}^*) - \varphi_2(\boldsymbol{x}_1)]$$

in bezug auf das Maß μ verteilt ist. Es ist nämlich $u(\boldsymbol{x}_1) \geqq 0$ für $\boldsymbol{x}_1 \in \mathfrak{X}$ und $\int u(\boldsymbol{x})\,\mu(d\boldsymbol{x}) = 1$. Dann gilt aber

$$\mathsf{M}\varkappa_2[f] = \int_{\mathfrak{X}} \frac{\mu(\mathfrak{X})}{\varphi_2(\boldsymbol{x}^*)\,\mu(\mathfrak{X})}\,[\varphi_2(\boldsymbol{x}^*)\,f(\boldsymbol{x}_1) - \varphi_2(\boldsymbol{x}_1)\,f(\boldsymbol{x}^*)]\,\mu(d\boldsymbol{x})$$
$$= \int_{\mathfrak{X}} f(\boldsymbol{x})\,\mu(d\boldsymbol{x})\,.$$

So haben wir eine Quadraturformel erhalten, deren eine Stützstelle fest ist und deren andere zufällig gewählt wird. Man kann das als eine gewisse Kombination der Methode der wesentlichen Stichprobe mit der Methode der Abtrennung des Hauptteils betrachten. Der Hauptteil von $f(\boldsymbol{x})$ wird hier durch den Ausdruck $f(\boldsymbol{x}^*)\,\varphi_2(\boldsymbol{x})$ geschätzt. Die elementar durchführbare Analyse der Dispersion von $\varkappa_2[f]$ zeigt allerdings, daß für die erfolgreiche Anwendung der hergeleiteten Formel eine bestimmte Glattheit der Funktionen $f(\boldsymbol{x})$ und $\varphi_2(\boldsymbol{x})$ in der Umgebung des Supremums notwendig ist. Außerdem muß die Menge $\mathfrak{X}$ bestimmte topologische Eigenschaften besitzen, damit das Integral

$$\int_{\mathfrak{X}} \frac{\varphi_2^2(\boldsymbol{x})\,[f(\boldsymbol{x}) - f(\boldsymbol{x}^*)]^2}{\varphi_2(\boldsymbol{x}) - \varphi_2(\boldsymbol{x}^*)}\,\mu(d\boldsymbol{x})$$

einen endlichen Wert besitzt.

Verhältnismäßig einfach kann man auch eine Interpolationsquadraturformel mit n Stützstellen herleiten, wobei $n - 1$ Stützstellen fest sind und eine zufällig gewählt wird. Wir werden die entsprechenden Resultate ohne ausführliche Be-

weise angeben, da sie nur für spezielle Problemstellungen von Interesse sind. Eine ausführliche Darstellung findet man in der Arbeit von GRANOWSKI und ERMAKOW [2].

Es seien $\boldsymbol{x}_1$ die stochastische Stützstelle und $\boldsymbol{x}_i = \boldsymbol{x}_i^*$, $i = 2, \ldots, n$, die festen Stützstellen. Wir betrachten die Interpolationsnäherungssumme $\varkappa_n[f] = \Delta(f, Q)/\Delta(Q)$ mit $Q = (\boldsymbol{x}_1, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^*)$ als Schätzwert für das Integral $\int \varphi_1(\boldsymbol{x}) f(\boldsymbol{x}) \mu(d\boldsymbol{x})$. Der Koeffizient $A_1(Q) = \Delta_1^{(1)}(Q)/\Delta(Q)$ entspricht dem Funktionswert im Punkt $\boldsymbol{x}_1$, die übrigen Koeffizienten $A_i^*(Q) = \Delta_i^{(1)}(Q)/\Delta(Q)$ den Funktionswerten in den festen Punkten. Wie schon früher bezeichnet $\Delta_i^{(1)}$ hier das algebraische Komplement des Elements $\varphi_1(\boldsymbol{x}_i)$ der Determinante Δ. Wir nehmen außerdem an, daß $\mathfrak{X}$ ein Gebiet $\mathcal{D}$ des s-dimensionalen euklidischen Raumes und μ das Lebesguesche Maß ist.

Mit denselben Methoden, wie wir sie bei der Untersuchung von stochastischen Quadraturformeln mit einer freien Stützstelle angewendet haben, kann man eine Formel für die Verteilungsdichte der Stützstelle $\boldsymbol{x}_1$ erhalten. Damit $\varkappa_n[f]$ eine erwartungstreue Schätzung liefert, muß $\boldsymbol{x}_1$ die Verteilungsdichte

$$u(\boldsymbol{x}_1) = \frac{\Delta(Q)}{\Delta_1^{(1)}(Q)} \varphi_1(\boldsymbol{x}_1) \tag{4.5.2}$$

besitzen. Die Funktion $u(\boldsymbol{x}_1)$ muß nichtnegativ sein, was bestimmte Einschränkungen für die Lage der festen Stützstellen nach sich zieht. Die Normierung von $u(\boldsymbol{x}_1)$ wird recht einfach nachgeprüft. Da $\Delta_1^{(1)}$ nicht von $\boldsymbol{x}_1$ abhängt, gilt

$$\int_{\mathcal{D}} u(\boldsymbol{x}_1)\, d\boldsymbol{x}_1 = \frac{1}{\Delta_i^{(1)}} \int \varphi_1(\boldsymbol{x}_1)\, \Delta(Q)\, d\boldsymbol{x}_1 = 1$$

wegen der Orthonormiertheit der Funktionen φ_i. Analog ergibt sich dann

$$\mathsf{M} A_i^*(Q) = \int_{\mathcal{D}} \frac{\Delta_i^{(1)}}{\Delta} \cdot \frac{\Delta}{\Delta_i^{(1)}} \varphi_1(\boldsymbol{x}_1)\, d\boldsymbol{x}_1 = \frac{1}{\Delta_i^{(1)}} \int_{\mathcal{D}} \Delta_i^{(1)} \varphi_1(\boldsymbol{x}_1)\, d\boldsymbol{x}_1 = 0\,,$$

$$\begin{aligned} \mathsf{M}\varkappa_n[f] &= \mathsf{M}(A_1(Q)\, f(\boldsymbol{x}_1)) \\ &= \int_{\mathcal{D}} \frac{\Delta_1^{(1)}}{\Delta} \cdot \frac{\Delta}{\Delta_1^{(1)}} \varphi_1(\boldsymbol{x}_1)\, f(\boldsymbol{x}_1)\, d\boldsymbol{x}_1 = \int_{\mathcal{D}} \varphi_1(\boldsymbol{x})\, f(\boldsymbol{x})\, d\boldsymbol{x}\,. \end{aligned}$$

Am lästigsten ist hierbei die Nichtnegativitätsbedingung für die durch die Gleichung (4.5.2) bestimmte Funktion u. Weil bei festen $\boldsymbol{x}_i^*$ der Nenner des Ausdrucks (4.5.2) eine Konstante ist, müssen die $\boldsymbol{x}_i^*$ so gewählt werden, daß der Ausdruck

$$\varphi_1(\boldsymbol{x}_1)\, \Delta(\boldsymbol{x}_1, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^*) \tag{4.5.3}$$

im Gebiet $\mathcal{D}$ sein Vorzeichen nicht ändert. Ist $\varphi_1(x) = \text{const}$ und sind die $\varphi_i(x)$, $i = 2, \ldots, n$, Polynome, dann kann man leicht zeigen, daß (4.5.3) nur dann sein Vorzeichen in einem gewissen Gebiet nicht ändert, wenn alle $\boldsymbol{x}_i^*$ auf dem Rand von $\mathcal{D}$ liegen. Einige hinreichende Bedingungen, unter denen das Vorzeichen von (4.5.3) konstant bleibt, werden in der Arbeit von GRANOWSKI und ERMAKOW [2] angegeben. Für ein Intervall lassen sich leicht Beispiele der genannten Quadratur-

formeln angeben, die für Polynome exakt gelten. Beispielsweise ist die Formel

$$\int_{-1}^{+1} f(x)\,dx \approx \frac{4}{3}(1-x_1^2)\,f(x_1) + \frac{1+3x_1}{3(1+x_1)}\,f(-1) + \frac{1-3x_1}{3(1-x_1)}\,f(+1)\,,$$

$$u(x_1) = \frac{3}{4}(1-x_1^2)\,, \qquad x_1 \in [-1, +1]\,,$$

exakt für Polynome zweiten Grades. Diese Formel wurde erstmalig von HAMMERSLEY und HANDSCOMB [1] erhalten.

Wir werden auch noch einige Quadraturformeln des genannten Typs im Fall mehrerer Veränderlicher angeben, wobei die ganze Zahl m bedeutet, daß die Quadraturformel für jedes Polynom vom Grade r mit $r \leqq m$ exakt gilt. In allen diesen Formeln könnten die festen Stützstellen auf dem Rand des Gebietes $\mathcal{D}$ auch anders gewählt werden. Für $s = 2$ und $m = 1$ gilt beispielsweise

$$\int_{-1}^{+1}\int_{-1}^{+1} f(x, y)\,dx\,dy \approx \frac{2}{1+x_1}\,[2f(x_1, y_1) - (y_1 - x_1)\,f(-1, +1) + (y_1 + x_1)\,f(-1, -1)]\,, \tag{4.5.4}$$

$$u(x_1, y_1) = \frac{1+x_1}{4}\,.$$

Es bereitet keine Schwierigkeiten, zu (4.5.4) analoge Formeln für eine größere Anzahl von Variablen herzuleiten. Die Anzahl der benötigten festen Stützstellen ist für $m = 1$ gleich s und wächst also entsprechend an. So erhalten wir beispielsweise für $s = 3$, $m = 1$

$$\int_{-1}^{+1}\int_{-1}^{+1}\int_{-1}^{+1} f(x, y, z)\,dx\,dy\,dz \approx \frac{4}{1+x_1}\,[2f(x_1, y_1, z_1) - 2(z_1 - x_1)\,f(-1, 0, 0) + (z_1 + y_1)\,f(-1, -1, -1) + (z_1 - y_1)\,f(-1, +1, -1)]\,,$$

$$u(x_1, y_1, z_1) = \frac{1+x_1}{8}\,.$$

Für $s = 2$, $m = 2$ hat eine Formel mit fünf festen Stützstellen auf dem Rand des Quadrats die Gestalt

$$\int_{-1}^{+1}\int_{-1}^{+1} f(x, y)\,dx\,dy$$

$$\approx \frac{1}{3(1-x_1^2)}\,\{8f(x_1, y_1) - [2y_1(1 - x_1 + 2y_1) - (1 + x_1)^2]\,f(-1, +1) + [2y_1(1 - x_1 - 2y_1) + (1 + x_1)^2]\,f(-1, -1) - 8(x_1^2 - y_1^2)\,f(-1, 0) + (1 + x_1)\,[(1 - 3x_1 - 2y_1)\,f(+1, +1) + (1 - 3x_1 + 2y_1)\,f(+1, -1)]\}$$

$$u(x_1, y_1) = \frac{3(1-x_1^2)}{8}\,.$$

Diese Formeln sind allerdings praktisch weniger wichtig mit Ausnahme der Fälle, in denen der Integrand eine besonders einfache Gestalt auf dem Rand besitzt (etwa gleich Null ist).

Man kann auch das Verfahren der voneinander abhängigen Stützstellen mit dem Verfahren der festen Stützstellen kombinieren. So erhält man beispielsweise für $s = 1$, $m = 5$ die Formel

$$\int_{-1}^{+1} f(x)\,dx \approx \frac{2}{15x_1^2(1-x_1^2)}[f(x_1)+f(-x_1)] + \frac{3-5x_1^2}{15(1-x_1^2)}[f(-1)+f(+1)] + \frac{20x_1^2-4}{15x_1^2}f(0)\,,$$

$$u(x_1) = \frac{15x_1^2(1-x_1^2)}{4}\,.$$

Quadraturformeln, bei denen Stützstellen von mehr als einer freien Stützstelle abhängen, haben wir nicht untersucht. Die einzige Ausnahme soll eine Formel mit zwei freien Stützstellen bilden

$$\int_{\mathcal{D}_s} f(\boldsymbol{x})\,d\boldsymbol{x} \approx \frac{1}{3}[f(\boldsymbol{x}_1)+f(\boldsymbol{x}_2)+f(\boldsymbol{x}_3)]\,. \tag{4.5.5}$$

Dabei ist $\mathcal{D}_s$ der s-dimensionale Einheitswürfel, $\{0 \leqq x^{(i)} \leqq 1,\ i = 1, \ldots, s\}$, die Koordinaten der Punkte $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ und $\boldsymbol{x}_3$ sind durch die Beziehung

$$(x_1^{(i)})^2 + (x_2^{(i)})^2 + (x_3^{(i)})^2 = 1\,, \qquad i = 1, \ldots, s\,,$$

miteinander verknüpft, $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ werden frei gewählt, und ihre Koordinaten besitzen die gemeinsame Verteilungsdichte

$$u(x_1, x_2) = \frac{2^s}{\pi^s}\prod_{k=1}^{s}\left(1-(x_1^{(k)})^2-(x_2^{(k)})^2\right)^{-1/2}. \tag{4.5.6}$$

Das entspricht einer Gleichverteilung der Größen $x_1^{(k)}$, $x_2^{(k)}$, $x_3^{(k)}$ auf einer Einheitskugelfläche.

Die Formel gilt exakt für Konstanten und die zweiten Potenzen von $x^{(k)}$. Wenn man symmetrisch zum Koordinatenursprung gelegene Stützstellen heranzieht, dann ergibt sich für den Würfel $\{-1 \leqq x^{(k)} \leqq 1,\ k = 1, \ldots, s\}$ eine Formel mit sechs Stützstellen, die für alle ersten und zweiten Potenzen von $x^{(k)}$ und für gewisse paarweise Produkte dieser Potenzen exakt gilt. Wir weisen darauf hin, daß die Anzahl der Stützstellen mit wachsendem s nicht größer wird. Die Dispersion der der Formel (4.5.5) entsprechenden Näherungssumme wurde von Ermakow [2] untersucht.

Die Stützstellen der symmetrischen Quadraturformel (4.5.5) wurden als Pseudozufallszahlen bei der Berechnung des Strahlendurchgangs durch Materie mit der Monte-Carlo-Methode verwendet. Zu diesem Zweck wurden Werte von s zwischen 50 und 100 gewählt. Die $6s$ Koordinaten der Stützstellen wurden in der Maschine gespeichert. Die Geschichte eines Teilchens wurde mit Hilfe von Zahlen nachgespielt, die den Koordinaten einer Stützstelle entsprechen. Reichten s Zahlen

für die Realisierung einer Geschichte nicht aus, so wurden weiter gewöhnliche Pseudozufallszahlen benutzt. Die Mittelwerte der entsprechenden Größe von sechs Teilchengeschichten bilden offensichtlich eine Realisierung der Näherungssumme.

Wenn man nebeneinander die Trajektorien von sechs Teilchen verfolgt, dann ist die vorherige Speicherung von Koordinaten der Stützstellen der Quadraturformel nicht nötig; man kann sie auch während der Teilchenbewegung erhalten, und die Größe s spielt keine wesentliche Rolle.

Die Berechnung wurde bei vorheriger Speicherung der Stützstellen durchgeführt. Es wurde der Durchgang von γ-Quanten durch eine Platte berechnet, wenn auf der einen Seite ein planparalleler Teilchenstrom eintritt.

Die Dispersion (über eine Trajektorie) der Austrittswahrscheinlichkeit der γ-Quanten aus der Platte auf der Seite, die dem einfallenden Strom gegenüber liegt, wird mit σ_0^2 bezeichnet, wenn die gewöhnliche Monte-Carlo-Methode benutzt wird, und mit σ_1^2 bei Gebrauch der Formel (4.5.5). Das Verhältnis von σ_0^2 zu σ_1^2 ist in der folgenden Tabelle angeführt (l bezeichnet dabei die Dicke der Platte in Vielfachen der freien Weglänge):

s	l					
	1	2	3	4	5	6
60	6,6	6,1	5,0	3,4	2,1	1,8
100	6,5	6,0	5,1	4,1	3,2	2,5

In den Rechnungen wurde bei der Modellierung der Verteilungen nicht die v.-Neumannsche Methode benutzt, da ihre Anwendung die Vielfachheit der zu berechnenden Integrale vergrößert.

2. Sukzessive Verfahren. Es wäre verlockend, ein Integrationsverfahren sukzessiver Art herzuleiten, in dem die in einem früheren Schritt erhaltenen Kenntnisse über die Funktion zur Verbesserung des Schätzwertes benutzt würden. Probleme solcher Art wurden beispielsweise von Marshall [1], Halton [2] und Pugh [1] untersucht. Die Idee, sukzessive Verfahren der Monte-Carlo-Methode für die Integralberechnung herzuleiten, hat gewisse Ähnlichkeit mit der Idee, bei der Methode der wesentlichen Stichproben (Kap. IV, § 2) Bernsteinpolynome zu benutzen.

Man kann die Rechnung in aufeinanderfolgende Schritte zerlegen und zugleich mit der Schätzung des Integrals $I_0 = \int f(\boldsymbol{x})\,\mu(d\boldsymbol{x})$ durch die Summe

$$\hat{c}_0 = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{x}_i)$$

im ersten Schritt die Fourierkoeffizienten

$$I_j = \int f(\boldsymbol{x})\,\varphi_j(\boldsymbol{x})\,\mu(d\boldsymbol{x}) \qquad \text{durch} \qquad \hat{c}_j = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{x}_i)\,\varphi_j(\boldsymbol{x}_i)$$

schätzen. Dabei sind die $\boldsymbol{x}_i$ unabhängig und entsprechend dem Wahrscheinlichkeitsmaß μ verteilt, die $\varphi_j \in \mathbf{L}_2(\mu)$ sind orthonormiert, $\varphi_0 \equiv 1$. Dann kann man annehmen, daß die Linearkombination

$$\hat{f}(\boldsymbol{x}) = \sum_{j=0}^{n} \hat{c}_j \varphi_j(\boldsymbol{x})$$

die Funktion f genügend gut approximiert. Wenn die entsprechenden Voraussetzungen erfüllt sind, kann man nun $\hat{f}$ bei der Anwendung der Methode der wesentlichen Stichprobe oder der Methode der Abtrennung des Hauptteils benutzen. In diesem zweiten Schritt werden dann auch die Größen I_j, $j = 0, \ldots, n$ und außerdem die Größen $I_{n+1}, \ldots, I_{n+m}$, $m > 0$, geschätzt usw.

Mit Hilfe der Ergebnisse von Kapitel V kann der Leser ohne Schwierigkeiten verschiedene Modifikationen sukzessiver Verfahren solcher Art herleiten. Im Zusammenhang mit diesen Verfahren stehen eine Reihe theoretischer Fragen, die bisher kaum untersucht wurden. Da die Anwendung dieser Verfahren die Rechenprogramme wesentlich kompliziert und es keine einfachen Empfehlungen für ihre Konstruktion gibt, wurde ihre Untersuchung nur in einer sehr geringen Anzahl von Fällen durchgeführt. Am vollständigsten wurden theoretische Fragen dazu von Halton [2] untersucht, wobei unter anderem auch folgendes Resultat erhalten wurde.

Es sei $\hat{f}_1, \hat{f}_2, \ldots$ eine Folge erwartungstreuer Schätzungen der Größe I_0. Wir nennen diese Schätzungen primär und stellen sekundäre Schätzungen $F_1, F_2, \ldots$ nach der Regel

$$F_m = a_m \hat{f}_m + (1 - a_m)\hat{f}_{m-1}, \qquad a_1 = 1, \qquad m = 1, 2, \ldots,$$

auf, so daß also

$$F_m = \sum_{i=1}^{m} b_i^{(m)} \hat{f}_i$$

mit

$$b_m^{(m)} = a_m, \quad b_i^{(m)} = a_i \prod_{r=i+1}^{m} (1 - a_r) \quad \text{für} \quad i < m, \qquad 0 \leqq b_i^{(m)} < 1,$$

gilt. Dann erhält man den

Satz 4.11. *Existiert* $\max_i b_i^{(m)} \to 0$ *und ein* $\varepsilon > 0$ *derart, daß die Reihe*

$$\sum_{k=1}^{\infty} a_k k^{-\varepsilon}$$

konvergiert und $\hat{\mathsf{D}} f_k \leqq c k^{-(1+2\varepsilon)}$ *mit einer Konstanten* c *gilt, dann konvergiert* F_m *für* $m \to \infty$ *mit der Wahrscheinlichkeit* 1 *gegen* I_0.

Zu den sukzessiven Prozeduren, die in der Literatur (Pugh [1]) betrachtet wurden, gehört auch die folgende.

Es sei $I_0 = \int f(\boldsymbol{x})\, \mu(d\boldsymbol{x})$ und $\nu(\boldsymbol{x}, \boldsymbol{y})$ eine vom Parameter $\boldsymbol{y}$ abhängige Schar von Wahrscheinlichkeitsmaßen, wobei $\boldsymbol{y} = (y_1, \ldots, y_n)$, $\boldsymbol{y} \in \mathcal{D}$, einem gewissen Gebiet $\mathcal{D}$ des n-dimensionalen euklidischen Raumes angehört. Das Maß μ soll für jedes $\boldsymbol{y} \in \mathcal{D}$ absolut stetig bezüglich ν sein.

Dann kann die Größe I_0 für jedes $\boldsymbol{y} = \boldsymbol{y}_0$ durch die Summe (Kap. IV, § 2)

$$\hat{I}_0(\boldsymbol{y}_0) = N^{-1} \sum_{i=1}^{N} f(\boldsymbol{x}_i^0) \frac{d\mu}{d\nu_0} (\boldsymbol{x}_i^0)$$

mit $\nu_0 = \nu(\boldsymbol{x}, \boldsymbol{y}_0)$ geschätzt werden. Dabei sind die $\boldsymbol{x}_i^0$ unabhängig für verschiedene i und in bezug auf das Maß ν_0 verteilt.

Die Dispersion des Schätzwertes $\hat{I}_0(\boldsymbol{y}_0)$ beträgt

$$\mathsf{D}\hat{I}_0(\boldsymbol{y}_0) = N^{-1} \left\{ \int f^2(\boldsymbol{x}) \left(\frac{d\mu}{d\nu_0} \right)^2 \nu_0(d\boldsymbol{x}) - I_0^2 \right\} \tag{4.5.7}$$

Man versucht nun, $\boldsymbol{y}_0$ so zu wählen, daß der Ausdruck (4.5.7) minimal wird. Das ist gleichbedeutend mit der Minimierung von

$$\psi(\boldsymbol{y}) = \int f^2(\boldsymbol{x}) \left(\frac{d\mu}{d\nu} (\boldsymbol{x}) \right)^2 \nu(d\boldsymbol{x}) \,.$$

Die Ermittlung eines solchen $\boldsymbol{y}$ kann mit einer Gradientenmethode erfolgen. Dazu wird eine Folge $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots$ und die entsprechende Folge von Maßen $\nu_0, \nu_1, \ldots$ folgendermaßen bestimmt:

$$\boldsymbol{y}_{k+1} = \boldsymbol{y}_k + \frac{h}{|\hat{\boldsymbol{v}}_{k+1}|} \hat{\boldsymbol{v}}_{k+1} \,.$$

$\hat{\boldsymbol{v}}_{k+1}$ ist ein Schätzwert des Gradienten $\boldsymbol{v}_{k+1}$ von ψ, also des Vektors mit den Komponenten

$$\boldsymbol{v}_{k+1} = \left(-\frac{\partial \psi}{\partial y_1}, \ldots, -\frac{\partial \psi}{\partial y_n} \right) \quad \text{für} \quad \boldsymbol{y} = \boldsymbol{y}_{k+1} \,,$$

h ist eine geeignet zu wählende Konstante (Schrittweite). Dabei muß man beachten, daß

$$\psi(\boldsymbol{y}_{k+1}) = \int f^2(\boldsymbol{x}) \left(\frac{d\mu}{d\nu_{k+1}} \right)^2 \nu_{k+1}(d\boldsymbol{x}) = \int f^2(\boldsymbol{x}) \frac{d\mu}{d\nu_{k+1}} \frac{d\mu}{d\nu_k} \nu_k(d\boldsymbol{x})$$

gilt und die Schätzung des Gradienten ψ mit Stichprobenwerten $\boldsymbol{x}$ durchgeführt werden kann, die entsprechend dem Maß ν_k verteilt sind. Pugh [1] gibt in seiner Arbeit ein numerisches Beispiel für die Anwendung der beschriebenen Methode auf eine Aufgabe der Massenbedienung an.

In Kapitel V werden wir das Problem betrachten, die Ableitungen des Erwartungswertes einer Zufallsfunktion auf Grund ihrer Stichprobenwerte in einzelnen Punkten zu schätzen. Die dort beschriebenen Algorithmen ermöglichen prinzipiell Die sukzessive Ermittlung solcher Parameter, die die Dispersion minimieren. Der Erfolg hängt allerdings wesentlich von der günstigen Wahl der Schar von Verteilungen ab und in vielen Fällen von der geeigneten Wahl der Anfangswerte, denn das Verhalten der Dispersion in Abhängigkeit von den Parametern kann sehr kompliziert sein.

V. DIE APPROXIMATION DER MITTELWERTE VON ZUFALLSFUNKTIONEN

§ 1. Gleichmäßige Approximation

In den §§ 1 und 2 dieses Kapitels werden die Ergebnisse der Arbeiten von FROLOW und TSCHENZOW [1], [2] dargestellt. Darin wird die Approximation von Funktionen untersucht, deren Werte mit Hilfe der Monte Carlo-Methode berechnet werden. Dabei wird, wie schon in Kapitel IV, nur die Aufgabe A betrachtet, es wird also die Endlichkeit des zweiten Moments der entsprechenden Zufallsgrößen vorausgesetzt. Die Untersuchung der Abhängigkeit eines stochastischen Modells von einem Parameter (oder von einer Gruppe von Parametern) ist für uns von besonderem Interesse. Sehr wichtig ist es, einen Parameterwert zu finden, für den ein gewisser Mittelwert ein Extremum erreicht. Von Interesse kann auch die Durchführung „vergleichender Experimente" sein, also die Lösung der Frage, welcher von zwei gewählten Werten des Parameters in bezug auf ein bestimmtes Kriterium das bessere Ergebnis liefert. Bei der Lösung dieser Probleme ist in vielen Fällen ein Verfahren bequem, das in der Arbeit von FROLOW und TSCHENZOW [1] dargestellt wird und das *Methode der abhängigen Versuche* genannt wurde. Obwohl eigentlich die abhängigen Versuche auch die Grundlage einiger anderer Methoden sind (Quadraturformeln mit stochastischen Stützstellen können abhängige Stützstellen haben), bleiben wir bei der Bezeichnung *Methode der abhängigen Versuche* für das anschließend darzustellende Verfahren, da diese Bezeichnung bereits in die Literatur eingegangen ist.

Es sei $\zeta(x, \omega)$ eine Funktion des zufälligen Arguments ω und des Parameters x. Wir bezeichnen dann den Erwartungswert von ζ bezüglich ω mit $f(x) = \mathsf{M}_\omega \zeta(x, \omega)$. Um die Werte von $f(x)$ über einem gewissen Intervall $[a, b]$ der Veränderlichen x zu erhalten, kann man eine Reihe von Werten x_i, $a \leqq x_1 < x_2 < \cdots < x_m \leqq b$, wählen und für diese x_i-Werte das Integral

$$\mathsf{M}_\omega \zeta = \int \zeta(x_i, \omega)\, \mathsf{P}(d\omega)\,, \qquad i = 1, \ldots, m\,,$$

mit der Monte-Carlo-Methode schätzen. Dabei werden wegen der statistischen Streuung die Werte des Integrals für verschiedene x_i im allgemeinen nicht auf einer glatten Kurve liegen, selbst dann wenn $f(x)$ eine glatte Funktion ist. Man kann verschiedene Glättungsmethoden anwenden, um eine glatte Kurve zu erhalten. Weiterhin werden wir solche Methoden ausführlich betrachten, die in der Regel durch die Berücksichtigung zusätzlicher Informationen über den Verlauf der Funktion $f(x)$ ihre Begründung finden. Das einfachste Verfahren besteht nun darin, bei der Schätzung des Integrals $\int \zeta(x_i, \omega)\, \mathsf{P}(d\omega)$ für verschiedene x_i ein und dieselbe Folge von Beobachtungen ω_j zu benutzen. In diesem Fall wird der Graph

$$N^{-1} \sum_{j=1}^{N} \zeta(x, \omega_j)$$

von sogar bei einer verhältnismäßig großen statistischen Streuung der Form nach dem Graph von $\mathsf{M}_\omega \zeta(x, \omega)$ benachbart sein, und eine Vorstellung vom Funktionsverlauf von $f(x)$ kann man mit einer kleinen Anzahl von Versuchen erhalten. Diese nur angedeuteten Überlegungen können auch streng begründet werden.

Wir bezeichnen nun mit $\varepsilon(x, \omega)$ die Differenz $\varepsilon(x, \omega) = \zeta(x, \omega) - f(x)$ und sagen, daß $\zeta(x, \omega) \in \Gamma_k$ ist, wenn die Bedingungen

$$\left|\frac{\partial^k \varepsilon(x, \omega)}{\partial x^k}\right| \leqq G_k(\omega) \tag{5.1.1}$$

und

$$\int [G_k(\omega)]^p \, \mathsf{P}(d\omega) = C_k < \infty \qquad (p \geqq 2)$$

erfüllt sind. Dann gilt der

Satz 5.1. *Gehört die Funktion $\zeta(x, \omega)$ zu $\bigcup_{k=1}^{l} \Gamma_k$, also zu jedem Γ_k, $k = 1, \dots, l$, dann konvergiert die zufällige Vektorfunktion*

$$\varphi_N = \sqrt{N} \left\{ N^{-1} \sum_{i=1}^{N} \frac{\partial^k(x, \omega_i)}{\partial x^k} - f^{(k)}(x) \right\}, \qquad k = 0, \dots, l-1 \tag{5.1.2}$$

gegen eine gewisse stetige zufällige Gaußsche Vektorfunktion.

Folgerung 5.1. *Sind für ζ die Bedingungen* (5.1.1) *für $k = 0, 1, \dots, l$ erfüllt, dann hat die Abweichung des empirischen Mittelwertes*

$$f_N^*(x) = N^{-1} \sum \zeta(x, \omega_i)$$

und seiner ersten $l - 1$ Ableitungen vom wahren Erwartungswert und seinen Ableitungen der Wahrscheinlichkeit nach eine Ordnung von $1/\sqrt{N}$.

Beweis des Satzes. Wegen

$$\varphi_N^{(k)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon^k(x, \omega_i)$$

folgt aus (5.1.2) die Ungleichung $\mathsf{D}\varphi_n^{(k)} \leqq C_k < \infty$. Nach dem Mittelwertsatz erhalten wir

$$\varphi_n^{(k)}(x'') - \varphi_n^{(k)}(x') = \varphi^{k+1}(\vartheta)\,(x'' - x')\,,$$

wobei ϑ zwischen x' und x'' liegt, und

$$\mathsf{M}|\varphi_n^{(k)}(x'') - \varphi_n^{(k)}(x')|^p \leqq |x'' - x'|^p \, C_{k+1}\,.$$

Daraus folgt, daß für $\varphi_n^{(k)}$, $k = 0, \dots, l-1$, gleichmäßig in k das Stetigkeitskriterium für Trajektorien von Kolmogoroff (Kap. I, § 2) erfüllt ist. Nach Prochorow [1] konvergiert eine stetige zufällige Vektorfunktion $\varphi_N(x)$ schwach gegen eine stetige zufällige Gaußsche Vektorfunktion $\overline{\varphi}_n(x)$, deren Erwartungswert identisch Null ist und deren Korrelationsfunktion

$$B_{j,k}(x', x'') = \mathsf{M}\,\{[\zeta^{(j)}(x', \omega) - f^{(j)}(x')]\,[\zeta^{(k)}(x'', \omega) - f^{(k)}(x'')]\}$$

ist. Damit ist der Satz bewiesen.

Das Maximum des Betrages der Vektorfunktion ist ein stetiges Funktional auf den Trajektorien des Prozesses und konvergiert folglich gegen eine gewisse Grenzverteilung. Wegen der Stetigkeit der Trajektorien des Grenzprozesses ist diese der Wahrscheinlichkeit nach beschränkt. Daraus folgt, daß die mit $\sqrt{N}$ multiplizierte Abweichung des empirischen Mittelwertes vom theoretischen Wert der Wahrscheinlichkeit nach beschränkt ist und folglich die Abweichung selbst der Wahrscheinlichkeit nach wie $1/\sqrt{N}$ abnimmt. Diese letzte Überlegung beweist die Folgerung des Satzes 5.1. Der bewiesene Satz gibt uns so die Möglichkeit, das unbekannte Bild der Funktion im Sinne der gleichmäßigen Konvergenz zu approximieren (eine Abnahmeordnung von $N^{-1/2}$ für das Maximum der absoluten Abweichung ist gesichert). Mit der Approximation im quadratischen Mittel beschäftigen wir uns in § 2.

Wenn das Integral $\int_{\mathcal{D}_{\boldsymbol{y}}} f(x, \boldsymbol{y})\, d\boldsymbol{y}$ zu berechnen ist, wobei $\mathcal{D}_{\boldsymbol{y}}$ ein (von x unabhängiges) Gebiet des s-dimensionalen euklidischen Raumes ist und $f(x, \boldsymbol{y})$ stetig differenzierbar von x abhängt, dann bedarf der Algorithmus keiner weiteren Erläuterung. Man muß für die Berechnung der Werte von $f(x, \boldsymbol{y})$ bei verschiedenen x dieselben Zufallszahlen benutzen.

Der Satz 5.1 bietet auch wesentliche Vorteile, wenn die Werte der Ableitung von $f(x)$ zu berechnen sind. Wir wollen die Ableitung $f'(x)$ im Punkt x_0 berechnen. Wäre die Funktion f exakt bekannt, so könnte man ein hinreichend kleines h wählen und (bei einer Rechnung mit der notwendigen Genauigkeit) den Wert von $f'(x_0)$ nach der Näherungsformel

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} \tag{5.1.3}$$

erhalten. Wir wollen nun diese Formel anwenden, indem wir $f(x_0 + h)$ und $f(x_0 - h)$ mit der Monte-Carlo-Methode durch zwei Serien ω_i, ω_i' unabhängiger Realisierungen der Größe ω schätzen.

$$\left.\begin{aligned} f(x_0 + h) &\approx N^{-1} \sum_{i=1}^{N} \zeta(x_0 + h, \omega_i) \\ f(x_0 - h) &\approx N^{-1} \sum_{i=1}^{N} \zeta(x_0 - h, \omega_i') \,. \end{aligned}\right\} \tag{5.1.4}$$

Dann wird sich der Fehler bei der Berechnung von $f'(x_0)$ aus dem Verfahrensfehler der Formel (5.1.3) und dem Fehler zusammensetzen, der bei der Schätzung der Funktionswerte mit den Formeln (5.1.4) entsteht. Für genügend kleines h kann man diesen Fehler also durch die Größe

$$h^2 |f'''(x_0)| + 3\gamma \frac{\sqrt{\mathcal{D}_0}}{\sqrt{2h^2 N}} \tag{5.1.5}$$

schätzen, wobei γ das Konfidenzniveau der Schätzung ist und

$$\mathcal{D}_0 = \int \zeta^2(x, \omega)\, \mathsf{P}(d\omega) - f^2(x_0)$$

gilt. Wenn h bei wachsendem N weniger als $N^{-1/2}$ abnimmt, strebt der Ausdruck (5.1.5) gegen Null. Man kann die optimale Darstellung für h angeben, die den Ausdruck (5.1.5) zum Minimum macht:

$$h_{\text{opt}} = (3\gamma)^{1/3} |f'''(x_0)|^{-1/3} [\mathsf{D}_0]^{1/6} (2N)^{-1/6} ;$$

das Minimum selbst ist gleich

$$2[9\gamma^2 |f'''(x_0)| \, \mathsf{D}_0]^{1/3} (2N)^{-1/3} .$$

Der Fehler nimmt also wie $N^{-1/3}$ ab.

Bei der Anwendung der Methode der abhängigen Versuche kann man dagegen die Schrittweite h beliebig klein wählen (bei entsprechender Rechengenauigkeit). Dann entfällt der Verfahrensfehler, und der Gesamtfehler ist gleich $\gamma \mathsf{D}_1^{1/2} N^{-1/2}$, wobei N die Anzahl der Versuche ist und

$$\mathsf{D}_1 = \int [\zeta'_x(x_0, \omega)]^2 \, \mathsf{P}(d\omega) - [f'(x_0)]^2 .$$

Analoge Berechnungen kann man bei komplizierteren Approximationsverfahren für die Ableitung und für Ableitungen höherer Ordnung durchführen. Bei der Approximation der ersten Ableitung durch kompliziertere Formeln und Anwendung der gewöhnlichen Monte-Carlo-Methode kann man im allgemeinen eine Fehlerabnahme von höherer Ordnung als $N^{-1/3}$ erreichen. Aber der Umfang der Rechenarbeit wächst dabei wesentlich an, und man muß eine größere Glattheit der Funktion fordern.

§ 2. Approximation einer unbekannten Dichte im quadratischen Mittel

In vielen Fällen möchte man die Dichte einer Zufallsgröße approximieren, deren Realisierungen mit der Monte-Carlo-Methode erhalten wurden. Eine Reihe von Aufgaben, in denen das benötigt wird, haben wir in Kapitel III betrachtet. Abschließend werden einige Fakten angeführt, die mit der einfachsten Approximation einer unbekannten Dichte im Mittel verbunden sind. Die Darstellung schließt sich an die Arbeit von Frolow und Tschenzow [1] an.

Es sei ξ eine Zufallsgröße und $\mu(dx)$ ein Maß im Raum der Werte von ξ. Wir bezeichnen mit $p(x) = d\mathsf{P}/d\mu$ die Verteilungsdichte von ξ in bezug auf das Maß μ und mit φ_i, $i = 1, \ldots, n$, ein System von mit dem Gewicht $r(x)$, $r(x) > 0$, orthonormierten Funktionen. Wir nehmen an, daß $p(x)$ dem Raum $\mathbf{L}_2(r, \mu)$ der mit dem Gewicht $r(x)$ bezüglich μ quadratisch integrierbaren Funktionen angehört. Wenn wir über Realisierungen der Zufallsgröße ξ verfügen, können wir die Fourierkoeffizienten (p, φ_i) der Funktion p bezüglich des Systems $\{\varphi_i\}$ schätzen. Man sieht leicht, daß $(p, \varphi_i) = \mathsf{M}[\varphi_i(\xi) \, r(\xi)]$ ist, und setzt dann wie gewöhnlich

$$(p, \varphi_i) \approx N^{-1} \sum_{j=1}^{n} [\varphi_i(\xi_j) \, r(\xi_j)] = a^*_{i,n} ,$$

wobei die ξ_i unabhängige Realisierungen von ξ sind. Mit den Schätzungen $a^*_{i,n}$ der Fourierkoeffizienten der Funktion $p(x)$ kann man das Polynom

$$\pi^*_m = \sum_{i=1}^{m} a^*_{i,n}\varphi_i(x)\sqrt{r(x)} \tag{5.2.1}$$

bilden, das als Schätzung für den Abschnitt $\pi_m(x)$ der Fourierreihe von $p(x)$ dient. Einige Aussagen über die Güte der Approximation von $p(x)$ durch $\pi^*_m(x)$ enthält der

Satz 5.2. *Gegeben seien eine Folge von Gewichten $r_m(x)$ und mit den Gewichten $r_m(x)$ orthonormierte Funktionensysteme $\{\varphi_{i,m}\}$, $i = 1, \dots, m$, mit*

$$||p(x)||^2 = \int p^2(x)\, r_m(x)\, \mu(dx) \leqq H < \infty\,. \tag{5.2.2}$$

Dann gilt:

A. *Aus*

$$||p(x) - \pi_m(x)||^2 < C_1 m^{-\gamma}\,, \tag{5.2.3}$$

$$m^{-1}\int \sum_{i=1}^{m} [\varphi_{i,m}(x)\, r_m(x)]^2\, p(x)\, \mu(dx) < C_2 \tag{5.2.4}$$

folgt mit $m \approx n^\beta$, $\beta = 1/(1+\gamma)$ die Ungleichung

$$\mathsf{P}\left(||p(x) - \pi^*_m(x)|| \geqq \left(\frac{C_1 + 2C_2}{\delta}\right)^{1/2} n^{-(1-\beta)/2}\right) \leqq \delta\,,$$

d. h., $R(\delta) = ((C_1 + 2C_2)/\delta)^{1/2}\, n^{-(1-\beta)/2}$ ist eine Bereichsschätzung für die Norm der Differenz $||p(x) - \pi^*_m(x)||$, die dem Konfidenzniveau $1-\delta$, $0 < \delta < 1$ entspricht.

B. *Ist*

$$B_1 m^{-\gamma} < ||p(x) - \pi_m(x)||^2\,, \tag{5.2.5}$$

$$4B_2 < m^{-1}\int \sum_{i=1}^{m} [\varphi_{i,m}(x)\, r_m(x)]^2\, p(x)\, \mu(dx)\,, \tag{5.2.6}$$

$$m^{-2}\int \sum_{i,l=1}^{m} [\varphi_{i,m}(x)\, r_m(x)]^2\, [\varphi_{l,m}(x)\, r_m(x)]^2\, p(x)\, \mu(dx) < C_3\,, \tag{5.2.7}$$

dann gilt von einem gewissen n ab (wie auch m für alle genügend kleinen δ gewählt wird) für jede Bereichsschätzung

$$R(\delta) > Bn^{-(1-\beta)/2}\,,$$

d. h.

$$\mathsf{P}(||p(x) - \pi^*_m(x)|| > Bn^{-(1-\beta)/2}) \geqq \delta\,, \tag{5.2.8}$$

wobei B eine geeignete Kombination der Konstanten B_1 und B_2 ist. Dabei kann die Bedingung (5.2.7) *noch abgeschwächt werden, indem man C_3 durch $\tilde{C}_3 m^{1+\gamma}$ ersetzt.*

Beweis. Die Differenz $p(x) - \pi^*_m(x)$ kann in beliebiger Reihenfolge nach den Maßen $\mu(dx)$ und $\mathsf{P}(d\omega)$ integriert werden, weil sie nach diesen Maßen meßbar ist. Die Differenz besteht aus zwei Komponenten, einer Komponente, die durch

den Unterschied von $p(x)$ und $\pi_m(x)$ hervorgerufen wird, und einer anderen, die durch die zufällige Abweichung der $a^*_{i,m}$ von $a_{i,m} = (p, \varphi_{i,m})$ bedingt ist. Dann gilt

$$\begin{aligned}||p(x)-\pi^*_m(x)||^2 &= \int [p(x) - \pi^*_m(x)]^2\, r_m(x)\, \mu(dx) \\ &= \int [p(x)-\pi_m(x)]^2\, r_m(x)\, \mu(dx) + \int [\pi_m(x)-\pi^*_m(x)]^2\, r_m(x)\, \mu(dx) \\ &= ||p(x) - \pi_m(x)||^2 + \sum_{i=1}^{m} (a^*_{i,m} - a_{i,m})^2 ,\end{aligned}$$

weil $p(x) - \pi_m(x)$ zu jedem $\varphi_{i,m}$ orthogonal ist.

Nach den Voraussetzungen des Satzes ist $||p(x)-\pi_m(x)||$ von der Ordnung $m^{-\gamma}$. Wir setzen

$$\eta^2 = \sum_{i=1}^{m} (a^*_{i,m} - a_{i,m})^2 ,$$

$$\mathsf{M}\eta^2 = \sum_{i=1}^{m} \mathsf{D}a^*_{i,m} = n^{-1}\left\{\sum_{i=1}^{m} \mathsf{M}\big(\varphi^2_{i,m}(\xi)\, r^2_m(\xi)\big) - \sum_{i=1}^{m} (\varphi_{i,m}, r_m)^2\right\}.$$

Aus den Voraussetzungen des Satzes folgt, daß der erste Summand in den geschweiften Klammern eine Größe der Ordnung m ist. Aus (5.2.2) folgt die Beschränktheit des zweiten Summanden. Daher sind die Voraussetzungen (5.2.4) und (5.2.6) folgenden Aussagen gleichwertig:

$$\mathsf{M}\eta^2 = \mathsf{M} \sum_{i=1}^{m} (a^*_{i,m} - a_{i,m})^2 < C_2 \frac{m}{n}, \tag{5.2.9}$$

$$2B_2 \frac{m}{n} < \mathsf{M} \sum_{i=1}^{m} (a^*_{i,m} - a_{i,m})^2 < C_2 \frac{m}{n} \quad \text{für} \quad m < \frac{H}{2B_2}. \tag{5.2.10}$$

Wir beweisen nun den ersten Teil des Satzes:

$$\mathsf{M}||p(x) - \pi^*_m(x)||^2 = ||p(x) - \pi_m(x)||^2 + \mathsf{M}\eta^2 < C_1 m^{-\gamma} + C_2 \frac{m}{n}.$$

Setzt man $m = 1 + [n^\beta]$ mit $\beta = (1 + \gamma)^{-1}$, so erhält man

$$\mathsf{M}||p(x) - \pi^*_m(x)||^2 < (C_1 + 2C_2)\, n^{-1+\beta} .$$

Daraus ergibt sich mit der Tschebyscheffschen Ungleichung

$$\mathsf{P}\left\{||p(x) - \pi^*_m(x)|| \geqq \left(\frac{C_1 + 2C_2}{\delta}\right)^{1/2} n^{-(1-\beta)/2}\right\} \leqq \delta .$$

Zum Beweis des zweiten Teiles wird zunächst $\mathsf{M}\eta^4$ geschätzt. Nach einigen Umformungen ergibt sich

$$\mathsf{M}\eta^4 < n^{-1}m^2n^{-2}[C_3 + 4C_2\sqrt{3} + 6C_2^2] + 3\left(\frac{m}{n}\right)^2 C_2^2 ,$$

also

$$\mathsf{M} \sum_{i,l=1}^{m} (a^*_{i,m} - a_{i,m})^2\, (a^*_{l,m} - a_{l,m})^2 < C_4 \left(\frac{m}{n}\right)^2 \tag{5.2.11}$$

mit $C_4 = 9C_2^2 + 4C_2\sqrt{C_3} + C_3$.

Nun benutzen wir die Ungleichung[1])

$$\mathsf{P}(\eta^2 \geqq \vartheta \mathsf{M}\eta^2) \geqq [\mathsf{M}\eta^2]^2 \, [\mathsf{M}\eta^4]^{-1} \, (1 - \vartheta)^2 \,. \tag{5.2.12}$$

Wegen dieser Ungleichung ist η^2 mit einer von Null verschiedenen Wahrscheinlichkeit von derselben Größenordnung wie $\mathsf{M}\eta^2$. Mit $\vartheta = \frac{1}{2}$ erhält man aus (5.2.10) und (5.2.12)

$$\mathsf{P}\left(\eta^2 \geqq \frac{\mathsf{M}\eta^2}{2}\right) \geqq \frac{B_2^2}{C_4}.$$

Folglich gilt mit einer Wahrscheinlichkeit, die größer als B_2^2/C_4 ist,

$$\begin{aligned} \|p(x) - \pi_m^*(x)\|^2 &= \|p(x) - \pi_m(x)\|^2 + \eta^2 > B_1 m^{-\gamma} + B_2 \frac{m}{n} \\ &\geqq [\gamma B_1 (B_2)^\gamma]^\beta \, n^{-1+\beta} \frac{1+\gamma}{\gamma}, \end{aligned}$$

wobei das Gleichheitszeichen nur für $m = [\gamma B_1 n / B_2]^\beta$ erreicht wird. Daher ergibt sich für $\delta < B_2^2/C_4$

$$\mathsf{P}\{\|p(x) - \pi_m^*(x)\| > B n^{-(1-\beta)/2}\} > \delta \,,$$

also gilt für jede Bereichsschätzung, die dem Konfidenzniveau $1 - \delta$ entspricht, die Beziehung $R(\delta) > B n^{-(1-\beta)/2}$.

Folgerung 5.2. *Die Abweichung des Histogramms (Häufigkeitsdiagramm) der Zufallsgröße ξ, $a \leqq \xi \leqq b$, vom Graph der Dichte ist bei einer Klassenauswahl $m \sim n^{1/3}$ bestenfalls der Wahrscheinlichkeit nach von der Ordnung $n^{-1/3}$.*

Beweis. Wir nehmen eine äquivalente Klasseneinteilung mit der Klassenbreite $h = (b - a)/m$ an. In diesem Fall ist $r(x) \equiv 1$, $\mu(dx) = dx$,

$$\varphi_{i,m}(x) = \begin{cases} \sqrt{\dfrac{m}{b-a}} & \text{für} \quad x \in d(i) \\ 0 & \text{sonst} \end{cases}$$

mit

$$d(i) = [a + ih, a + (i+1)\,h], \quad i = 0, 1, \ldots, m-1 \,.$$

---

[1]) Diese Ungleichung ist eine Folgerung der Ungleichung

$$(\int |x| \, dG(x))^2 \leqq \int x^2 \, dG(x)$$

für die Momente. Es sei nämlich ξ eine Zufallsgröße und $\vartheta > 0$ fest. Dann erhält man mit

$$G(x) = \begin{cases} 0 & \text{für} \quad x < \vartheta \, \mathsf{M}|\xi| \,, \\ (\mathsf{P}(|\xi| > \vartheta \mathsf{M}|\xi|))^{-1} \, \mathsf{P}(|\xi| > x) & \text{für} \quad x \geqq \vartheta \, \mathsf{M}|\xi| \end{cases}$$

die Ungleichung

$$\left(\int_{\vartheta \mathsf{M}|\xi|}^{\infty} |x| \, d\mathsf{P}(|\xi| > x)\right)^2 \leqq \int_{\vartheta \mathsf{M}|\xi|}^{\infty} x^2 \, d\mathsf{P}(|\xi| > x) \, \mathsf{P}(|\xi| > \vartheta \mathsf{M}|\xi|) \,.$$

Die Ungleichung (5.2.12) ist eine Folgerung aus dieser letzten Ungleichung und den Ungleichungen

$$\int_{\vartheta \mathsf{M}|\xi|} x^2 \, d\mathsf{P}(|\xi| > x) \leqq \int_0^{\infty} x^2 \, d\mathsf{P}(|\xi| > x) = \mathsf{M}\xi^2 \,,$$

$$\int_{\vartheta \mathsf{M}|\xi|}^{\infty} |x| \, d\mathsf{P}(|\xi| > x) = \mathsf{M}|\xi| - \int_0^{\vartheta \mathsf{M}|\xi|} |x| \, d\mathsf{P}(|\xi| > x) \geqq \mathsf{M}|\xi| - \vartheta \mathsf{M}|\xi| = \mathsf{M}|\xi| \, (1 - \vartheta) \,.$$

Die Voraussetzungen (5.2.3), (5.2.6) und (5.2.7) des Satzes sind erfüllt, wie sich folgendermaßen ergibt:

$$\frac{1}{m}\int \sum_{i=1}^{m} [\varphi_{i,m}(x)]^2\, p(x)\, dx = \frac{1}{m}\int_a^b \frac{m}{b-a}\, p(x)\, dx = \frac{1}{b-a},$$

$$\frac{1}{m^2}\int \sum_{i,l=1}^{m} [\varphi_{i,m}(x)]^2\, [\varphi_{l,m}(x)]^2\, p(x)\, dx = \frac{1}{m^2}\int_a^b \frac{m^2}{(b-a)^2}\, p(x)\, dx = \frac{1}{(b-a)^2}.$$

Für $x \in d(i)$ ist $\pi_m(x) = h^{-1} \int_{d(i)} p(x)\, dx$. Setzt man die zweite Ableitung d^2p/dx^2 als beschränkt voraus, so ergibt sich

$$\int_a^b |p(x) - \pi_m(x)|^2\, dx = \frac{h^2}{12}\int_a^b \left[\frac{d^2p}{dx^2}\right]^2 dx + O(h^3) \sim m^{-2}.$$

Daraus folgt $\beta = 1/3$.

Nach dem bewiesenen Satz ist die zufällige Abweichung $||p(x) - \pi_m^*(x)||$ von einer Größenordnung nicht kleiner als $n^{-1/3}$. Diese Schranke wird für $m \sim n^{1/3}$ erreicht, und in anderen Fällen ist das Ergebnis nur schlechter.

Das letzte Ergebnis besitzt ein mehrdimensionales Analogon, das wir hier ohne Beweis formulieren.

Folgerung 5.3. *Die Abweichung des Histogramms eines Zufallsvektors* $\xi = \{\xi^{(j)}\}$, $a_j \leqq \xi^{(j)} \leqq b_j$, $j = 1, \ldots, s$, *vom Graph der Dichte ist bei äquivalenter Klasseneinteilung bezüglich jeder Koordinate und einer Klassenanzahl* $m \sim n^{s/(s+2)}$ *der Wahrscheinlichkeit nach von der Größenordnung* $n^{-1/(2+s)}$.

Folgerung 5.4. *Die Funktionen* φ_i *bilden ein vollständiges System mit dem Gewicht* $r(x)$ *orthonormierter Funktionen. Sind die Bedingungen*

$$|\int p(x)\, \varphi_i(x)\, r(x)\, \mu(dx)| < A i^{-\alpha}, \qquad \alpha > 1/2,$$

$$\int [\varphi_i(x)\, r(x)]^2\, p(x)\, \mu(dx) < C$$

erfüllt und wählt man $m = n^{1/2\alpha}$, *so ist die Abweichung des Polynoms* $\pi_m^*(x)$ *von* $p(x)$ *in der Norm des* $\mathbf{L}_2(r, \mu)$ *der Wahrscheinlichkeit nach von der Größenordnung* $n^{-1/2+1/4\alpha}$.

Beweis. Nach der Parsevalschen Gleichung gilt

$$||p(x) - \pi_m(x)||^2 = \sum_{i=m+1}^{\infty} (\varphi_i, p)^2 < A(2\alpha - 1)^{-1}\, m^{-2\alpha+1}.$$

Dann sind die Bedingungen des Satzes 5.2 mit $\gamma = 2\alpha - 1$, $C_1 = A/(2\alpha - 1)$ und $C_2 = C$ erfüllt, woraus die Behauptung folgt.

Der Satz 5.2 und seine Folgerungen ermöglichen die Darstellung der Berechnungen mit der Monte-Carlo-Methode in Form einer approximativen analytischen Abhängigkeit. Wesentlich ist dabei, daß die numerischen Algorithmen einfach sind. Die Anwendung der beschriebenen Methoden erfordert einige a-priori-

Kenntnisse über die Glattheit der zu approximierenden Funktionen. Manchmal kann es vorteilhaft sein, die ursprüngliche Aufgabe in eine vielleicht sogar kompliziertere Form zu transformieren, wenn dabei die uns interessierenden Abhängigkeiten glatte Funktionen der Parameter werden.

Die Schätzung der Genauigkeit bei der Berechnung der Ableitung durch unabhängige Versuchsreihen und bei der Approximation der Dichte durch ihre Fourierreihe stellt die Verbindung zwischen den zufälligen und systematischen Komponenten des Verfahrensfehlers her. Zugleich ist am Beispiel der Methode der abhängigen Versuche ersichtlich, daß spezielle Modellierungsmethoden zu einer wesentlichen Einsparung bei der Rechenarbeit führen können. Eine analoge Erscheinung wurde auch bei der Berechnung von Integralen festgestellt.

Der folgende Paragraph ist der Approximation einer Zufallsfunktion durch einen Abschnitt ihrer Fourierreihe in den Fällen gewidmet, in denen wir über Parameterwerte verfügen können, in welchen die Zufallsfunktion realisiert wird.

§ 3. Interpolation mit stochastischen Stützstellen

Die Approximationsmethode für Funktionen, die wir in diesem Paragraphen beschreiben werden, gründet sich formal auf die Methode der Interpolation. Sie hängt eng mit der Methode der kleinsten Quadrate und mit Aufgaben der Versuchsplanung bei einer Regressionsanalyse zusammen. Die anschließend dargestellten Resultate wurden hauptsächlich in den Arbeiten von Ermakow [1], [3] und Granowski, und Ermakow [1] erhalten.

Im vorigen Paragraphen haben wir für die Schätzung der Fourierkoeffizienten einer unbekannten Funktion die gewöhnliche Monte-Carlo-Methode benutzt. Jetzt wollen wir für die Schätzung der Fourierkoeffizienten Quadraturformeln mit stochastischen Stützstellen benutzen. Im wesentlichen löst die Einführung einer Funktion φ_1 in den Integranden dieses Problem (es wurde schon mehrfach bemerkt, daß anstelle von φ_1 auch eine andere Funktion φ_i genommen werden kann, für die die Quadraturformel exakt gilt). Daher könnten wir auf das vorige Kapitel verweisen, da die dort erzielten Ergebnisse bei der Berechnung der Fourierkoeffizienten der uns interessierenden Funktion anwendbar sind. Es gibt aber gewisse Besonderheiten, die für die Approximationsaufgabe charakteristisch sind und eine gesonderte Darstellung der betreffenden Resultate nützlich erscheinen lassen. Die Verbindungen zu Aufgaben der Versuchsplanung einer Regressionsanalyse ermöglichen auch die Betrachtung einiger Aspekte bei der Integralberechnung, die bisher noch nicht untersucht wurden.

Wie früher sei $\zeta(\boldsymbol{x}, \omega)$ eine Zufallsfunktion, $\boldsymbol{x} \in \mathfrak{X}$, ω eine Zufallsgröße, $\omega \in \Omega$. Wir nehmen noch an, daß $\mathsf{M}_\omega \zeta(\boldsymbol{x}, \omega) = f(\boldsymbol{x})$ ist. Dann wollen wir $f(\boldsymbol{x})$ durch eine Linearkombination von Funktionen φ_i, $i = 1, \ldots, n$ unter der Voraussetzung approximieren, daß Realisierungen der Funktion ζ für gewisse Werte des Parameters $\boldsymbol{x}$ bekannt sind. Die wesentliche Voraussetzung, die wir hier machen und die die Aufgaben dieses Paragraphen von denen des vorigen unterscheidet, besteht

darin, daß wir die Parameterwerte $\boldsymbol{x}$, für die die Berechnung (Messung) von $\zeta(\boldsymbol{x})$ durchgeführt werden soll, nach eigenem Belieben auswählen können. Das Problem einer in bezug auf ein gewisses Kriterium optimalen Wahl der Parameterwerte $\boldsymbol{x}$ ist dem Wesen nach eine Aufgabe der Versuchsplanung, wie sie auch in diesem Paragraphen betrachtet wird.

Wir nehmen nun an, daß ζ eine Hilbertsche Zufallsfunktion und $f \in \mathbf{L}_2(\mu)$ ist; die $\varphi_i(\boldsymbol{x}) \in \mathbf{L}_2(\mu)$, $i = 1, \ldots, n$ seien linear unabhängig; mit ζ_j bezeichnen wir die Realisierung der Funktion ζ für $\boldsymbol{x} = \boldsymbol{x}_j$. Dann betrachten wir folgende Interpolationsaufgabe.

Zu bestimmen sind solche Zahlen a_l, daß die Gleichungen

$$\sum_{l=1}^{n} a_l \varphi_l(\boldsymbol{x}_j) = \zeta_j \,, \qquad j = 1, \ldots, n \,, \tag{5.3.1}$$

gelten. Wir setzen nun voraus, daß die Determinante $\Delta(Q)$ dieses Systems ungleich Null ist, so daß die a_l berechnet werden können. Unter Benutzung der Cramerschen Regel erhalten wir die folgende bekannte Form des Interpolationspolynoms:

$$\mathcal{L}[\zeta, Q] = -\frac{1}{\Delta(Q)} \begin{vmatrix} 0 & \varphi_1(\boldsymbol{x}) & \ldots & \varphi_n(\boldsymbol{x}) \\ \zeta_1 & \varphi_1(\boldsymbol{x}_1) & \ldots & \varphi_n(\boldsymbol{x}_1) \\ \ldots & \ldots & \ldots & \ldots \\ \zeta_n & \varphi_1(\boldsymbol{x}_n) & \ldots & \varphi_n(\boldsymbol{x}_n) \end{vmatrix} . \tag{5.3.2}$$

Die a_l selbst sind gleich

$$a_l = \frac{(-1)^l}{\Delta(Q)} \det \|\zeta_i, \varphi_1(\boldsymbol{x}_i), \ldots, \varphi_{l-1}(\boldsymbol{x}_i), \varphi_{l+1}(\boldsymbol{x}_i), \ldots, \varphi_n(\boldsymbol{x}_i)\|, \qquad l = 1, \ldots, n. \tag{5.3.3}$$

Man sieht leicht, daß $a_1 = \varkappa_n[f]$ ist (in den Bezeichnungen von § 2) und daß die übrigen a_l in $\varkappa_n[f]$ übergehen, wenn man dort anstelle von φ_l die Funktion φ_1 einsetzt und die Spalten entsprechend vertauscht. Aus (5.3.2) folgt auch, daß wir die φ_i ohne Beschränkung der Allgemeinheit als orthonormiert voraussetzen können. Die Bildung der gleichen Linearkombinationen von Spalten in Zähler und Nenner von (5.3.2) ändert die Größe $\mathcal{L}[\zeta, Q]$ nicht. Weil wir die Werte $\boldsymbol{x}_j$ frei und insbesondere auch zufällig wählen können, gilt der

Satz 5.3. *Bilden die φ_i, $i = 1, \ldots, n$, ein orthonormiertes Funktionensystem und ist die Zufallsgröße Q mit der Dichte $w_0(Q) = (n!)^{-1} \Delta^2(Q)$ verteilt, so ist der totale Erwartungswert von $\mathcal{L}[\zeta, Q]$ gleich der Teilsumme der Fourierreihe der Funktion f nach dem System $\{\varphi_i\}$*

$$\mathsf{M}_{\omega, Q} \mathcal{L}[\zeta, Q] = \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x}) \,. \tag{5.3.4}$$

Hierin bezeichnet ω die Gesamtheit der Werte $\omega_1, \ldots, \omega_n$ für die Realisierungen der Funktion ζ erhalten wurden.

Der Beweis des Satzes folgt aus den vorhergehenden Überlegungen und aus Satz 4.3.

In analoger Weise folgt aus Satz 4.4 der

Satz 5.4. *Die Dispersion der Zufallsgröße $\mathcal{L}[\zeta, Q]$ genügt der Ungleichung*

$$\mathsf{D}_{\omega,Q}\mathcal{L}[\zeta, Q] \leqq \sum_{i=1}^{n} \varphi_i^2(\boldsymbol{x}) \left\{\int [f[\boldsymbol{x}] - \sum_{j=1}^{n} (\varphi_j, f)\, \varphi_j(\boldsymbol{x})]^2 \mu(d\boldsymbol{x}) + \int \sigma^2(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right\}, \tag{5.3.5}$$

wobei für ein reguläres System $\{\varphi_i\}$ das Gleichheitszeichen gilt.

Der Fall, daß die Realisierungen $\zeta(\boldsymbol{x}_j, \omega)$ für nur einen Wert erhalten wurden (der Fall abhängiger Versuche), wird von uns nicht betrachtet. Man kann zeigen, daß die Dispersion des Interpolationspolynoms in diesem Fall kleiner wird.

Der beschriebene Algorithmus für die Interpolation ist also äquivalent der gleichzeitigen Berechnung von n Fourierkoeffizienten der Funktion f. Wir nehmen nun an, daß wir zur Verwirklichung des Algorithmus die Werte von Q mit der v.-Neumannschen Methode bestimmen. Dazu sei $\mathfrak{X}$ ein Gebiet $\mathcal{D}$ des s-dimensionalen Raumes mit dem Volumen 1 und μ das Lebesguesche Maß. Dann besteht der Algorithmus in folgendem:

1. Man wählt die im Gebiet $\mathcal{D}$ gleich verteilten Größen $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ und berechnet $\Delta(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$, indem man beispielsweise die Matrix $\varphi_i(\boldsymbol{\alpha}_j)$ in zwei Dreiecksmatrizen zerlegt. Wenn $|\Delta(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)| > \alpha M$ ausfällt, wobei α nicht von den $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ abhängt, und M das Maximum des Betrages der Determinante $\Delta(Q)$ ist, dann kann man $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ als Realisierung für Q wählen. Im entgegengesetzten Fall muß die Rechnung mit neuen $\boldsymbol{\alpha}_j$ und α wiederholt werden.

2. Nachdem ein geeignetes Q gefunden wurde, werden nun die ζ_j berechnet. Mit Hilfe der Zerlegung der Matrix $\varphi_i(\boldsymbol{\alpha}_j)$ in zwei Dreiecksmatrizen ist das System

$$\sum_{i=1}^{n} \varphi_i(\boldsymbol{\alpha}_j)\, a_i = \zeta_j\,, \qquad j = 1, \ldots, n\,,$$

leicht zu lösen. Dabei müssen die φ_i nicht unbedingt orthonormiert sein. Der Mittelwert der Linearkombination $\sum_{i=1}^{n} a_i\varphi_i(\boldsymbol{x})$ stellt jedenfalls eine Teilsumme der Fourierreihe von f nach dem System dar, das durch Orthonormierung der Funktionen φ_i entstanden ist. Eine andere interessante Besonderheit besteht darin, daß wegen der Struktur der Dispersion (5.3.5) leicht Hypothesen über die Anpassung an die gesuchte Funktion geprüft werden können. Diese Behauptung soll ausführlicher erklärt werden.

Im Sinne des dritten Kapitels sei ein Modell gegeben, das von dem Parameter x abhängt. Das soll beispielsweise das Modell des Strahlendurchgangs durch eine Schicht der Dicke H sein. Wir nehmen an, daß die x-Achse senkrecht auf der Schicht steht. Die Lage der punktförmigen Quelle auf der x-Achse ist der Parameter (beschrieben durch die Koordinate x). Wir setzen $0 \leq x \leq H$ voraus. Wenn uns die Wahrscheinlichkeit des Strahlenaustritts aus der Schicht bei $x = H$ interessiert und wir diese mit der Monte-Carlo-Methode schätzen, dann kann das Ergebnis der Berechnungen von N Realisierungen bei fester Lage x der Quelle als Realisierung einer gewissen Zufallsfunktion $\zeta_N(x, \omega)$ angesehen werden. Nun wollen wir die wahre Abhängigkeit $f(x)$, die die Wahrscheinlichkeit des Strahlenausgangs aus der Schicht beschreibt, durch eine Linearkombination $\sum_{i=1}^{n} a_i\varphi_i(x)$

der Funktionen φ_i approximieren (die φ_i seien in diesem Fall orthonormiert). Nachdem wir mit dem beschriebenen Algorithmus einen Wert $Q = (x_1, ..., x_n)$ erhalten haben, berechnen wir $\zeta_N(x_i, \omega)$, $i = 1, ..., n$, und bestimmen die Koeffizienten a_i. Die Prozedur kann mehrfach durchgeführt werden (N braucht nicht allzu groß zu sein). Zugleich kann die Dispersion der a_i geschätzt werden. Diese Dispersion besteht aus zwei Summanden:

$$N^{-1} \int_0^H \left[f(x) - \sum_{j=1}^n (\varphi_j, f)\, \varphi_j(x)\right]^2 dx \quad \text{und} \quad N^{-1} \int_0^H \sigma^2(x)\, dx\,.$$

Werden alle Rechnungen für ein und dieselben Q-Werte durchgeführt, aber beispielsweise mit verdoppelter Anzahl von Realisierungen für jedes Q (es wird $\zeta_{2N}(x, \omega)$ berechnet), dann verringert sich der zweite Summand auf die Hälfte. So ist es möglich, die Summanden zu trennen und außerdem zu bestimmen, um wieviel genauer die gewählte Linearkombination die uns interessierende Funktion $f(x)$ approximiert. Bei vielen physikalischen Aufgaben, bei denen man von vornherein Aussagen über die interessierende Funktion machen kann, wird ein solches Verfahren gute Dienste leisten.

Der Begriff der Zulässigkeit, der in Verbindung mit den Integrationsverfahren eingeführt wurde, kann in natürlicher Weise auch für die Verfahren der stochastischen Interpolation erklärt werden. Hier gibt es allerdings eine größere Willkür, weil eine Vektorgröße (die Gesamtheit der Koeffizienten a_i) geschätzt wird, und man kann verschiedene Kriterien für die Güte der Schätzung betrachten. Auf eines davon gehen wir näher ein.

Wie vorher halten wir $\mathcal{L}[\zeta, Q]$ fest und betrachten dann alle möglichen Verteilungsfunktionen $W(Q)$ der Zufallsgröße Q. Die Gesamtheit $\pi = (\mathcal{L}[\zeta, Q], W(Q))$ nennen wir eine Prozedur der zufälligen Interpolation, wenn

$$\mathsf{M}\mathcal{L}[\zeta, Q] = \sum_{i=1}^n (\varphi_i, f)\, \varphi_i(x) \quad \text{und} \quad \mathsf{D}\mathcal{L}[\zeta, Q] < \infty$$

ist. Die Verteilungsfunktion $W(Q)$, für die (5.3.4) erfüllt ist, nennt man natürlicherweise auch einen stochastischen Versuchsplan (aus Gründen, auf die wir später eingehen).

Wir werden sagen, daß $\tilde{\pi} = (\tilde{\mathcal{L}}[\zeta, Q], \tilde{W})$ in einer gegebenen Klasse von Prozeduren $\mathcal{P}$ für eine Funktionenklasse $\mathbf{H}$ über π dominiert (vorausgesetzt, daß $\mathcal{L}[\zeta, Q]$ in der Klasse $\mathcal{P}$ die Gestalt $\sum_{i=1}^n a_i \varphi_i(\boldsymbol{x})$ hat), wenn $\mathsf{D}_{Q_{\tilde{W}}} \tilde{a}_j \leqq \mathsf{D}_{Q_W} a_j$ für beliebige Funktionen $\zeta \in \mathbf{H}$ und $j = 1, ..., n$ gilt, wobei wenigstens für ein $\zeta = \zeta_0 \in \mathbf{H}$ und einen Werte $j = j_0$ die strenge Ungleichheit gilt. π heißt zulässig in der Klasse $\mathcal{P}$ für $\mathbf{H}$, wenn es in $\mathcal{P}$ für $\mathbf{H}$ keine über π dominierende Prozedur gibt. Bei dieser Definition des Zulässigkeitsbegriffes gelten Sätze, die den Sätzen 4.5 und 4.6 analog sind (Kap. IV, § 3).

Satz 5.5. *Ist das System $\{\varphi_i\}$ in bezug auf das Maß μ regulär und bilden die Funktionen $\{\varphi_i(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})\}$ $(i, j = 1, ..., n)$ kein vollständiges System in $\mathbf{L}_2(\mu)$, dann ist die Prozedur $(\mathcal{L}[f, Q], W_0)$ für die Klasse $\mathbf{H}$ der Hilbertschen Zufallsfunktionen nicht zulässig.*

Satz 5.6. *Für das System der Haarschen Funktionen ist jede Prozedur der stochastischen Interpolation* $(\mathcal{L}[f, Q], W)$ *für* $\mathbf{H}$ *zulässig.*

Die Beweise dieser Sätze sind analog den Beweisen der Sätze 4.5 bzw. 4.6, so daß wir sie nicht durchführen werden.

Wir gehen nun auf den Zusammenhang zwischen einer stochastischen Interpolationsformel und der Methode der kleinsten Quadrate ein. Dazu benötigen wir die bekannte Formel von Binet-Cauchy, die einen Sonderfall der Aussage von Lemma 4.1 darstellt. Es seien nämlich φ_k, $k = 1, \ldots, n$, und ψ_l, $l = 1, \ldots, n$, zwei Systeme linear unabhängiger Funktionen auf dem Träger des Maßes μ. Dabei ist μ ein auf der endlichen Punktmenge $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ konzentriertes Maß, das gleich p_i, $i = 1, \ldots, N$, für jeden Punkt ist. Die Anwendung von Lemma 4.1 führt in diesem Fall zu der Gleichung

$$n! \det \left\| \sum_{i=1}^{N} p_i \varphi_k(\boldsymbol{x}_i)\, \psi_l(\boldsymbol{x}_i) \right\| = \sum_{i_1=1}^{N} p_{i_1} \cdots \sum_{i_n=1}^{N} p_{i_n} \det \|\varphi_k(\boldsymbol{x}_{i_p})\| \det \|\psi_l(\boldsymbol{x}_{i_q})\| .$$

Dabei ist zu berücksichtigen, daß beim Zusammenfallen von Werten i_q, $q = 1, \ldots, n$, die entsprechenden Determinanten verschwinden. Das ergibt

$$n! \det \left\| \sum_{i=1}^{N} p_i \varphi_k(\boldsymbol{x}_i)\, \psi_l(\boldsymbol{x}_i) \right\| = \sum_{i_1 \neq i_2 \neq \cdots \neq i_n} p_{i_1} \cdots p_{i_n} \det \|\varphi_k(\boldsymbol{x}_{i_p})\| \det \|\psi_l(\boldsymbol{x}_{i_q})\| .$$

Berücksichtigt man jetzt, daß die Vertauschung jeweils derselben Spalten (Zeilen) in den unter dem Summenzeichen stehenden Determinanten ihr Produkt nicht ändert, so erhält man

$$\begin{aligned} &n! \det \left\| \sum_{i=1}^{n} p_i \varphi_k(\boldsymbol{x}_i)\, \psi_l(\boldsymbol{x}_i) \right\| \\ &= n! \sum_{1 \leq i_1 < \cdots < i_n \leq N} p_{i_1} \cdots p_{i_n} \det \|\varphi_k(\boldsymbol{x}_{i_p})\| \det \|\psi_l(\boldsymbol{x}_{i_q})\| . \end{aligned} \tag{5.3.6}$$

Für $p_i = 1$, $i = 1, \ldots, N$, fällt die Formel (5.3.6) mit der Formel von Binet-Cauchy für die Determinante des Produktes zweier rechteckiger Matrizen zusammen.

Anstelle eines Interpolationsverfahrens zur Approximation der Funktion $f(\boldsymbol{x})$ soll jetzt die Methode der kleinsten Quadrate benutzt werden, d. h., aus den Realisierungen der Funktion $\zeta(\boldsymbol{x}, \omega)$ in N Punkten (der Einfachheit halber seien die Realisierungen unabhängig) wird eine Approximation der Funktion $f(\boldsymbol{x})$ in Form der Linearkombination $\sum_{i=1}^{n} a_i \varphi_i(\boldsymbol{x})$, $N > n$, durch die Bedingung

$$\sum_{j=1}^{N} p_j \left[\zeta_j - \sum_{q=1}^{n} a_q \varphi_q(\boldsymbol{x}_j) \right]^2 = \min \tag{5.3.7}$$

ermittelt. Für $p_j = 1$ werden die a_q bekanntlich (Linnik [1]) durch die Formeln

$$a_q = (-1)^{q+1} \frac{\det \|[\zeta, \varphi_i], [\varphi_2, \varphi_i], \ldots, [\varphi_{q-1}, \varphi_i], [\varphi_{q+1}, \varphi_i], \ldots, [\varphi_n, \varphi_i]\|}{\det \|[\varphi_1, \varphi_i], [\varphi_2, \varphi_i], \ldots, [\varphi_n, \varphi_i]\|} \tag{5.3.8}$$

mit $[\varphi, \psi] = \sum_{i=1}^{N} \varphi(\boldsymbol{x}_i)\, \psi(\boldsymbol{x}_i)$ bestimmt. Unter Anwendung der Formel (5.3.6) erhält man dann folgendes Resultat:

Die sich aus der Bedingung (5.3.7) ergebenden Größen a_q sind die Mittelwerte der Größen $a_q(j_1, \ldots, j_n)$, die den Gleichungen

$$\sum_{q=1}^{n} a_q(j_1, \ldots, j_n)\, \varphi_q(\boldsymbol{x}_{j_k}) = \zeta_{j_k}, \qquad k = 1, \ldots, n,$$

für alle möglichen $j_1, \ldots, j_n$ genügen. Dabei wird die Mittelbildung mit dem Gewicht

$$\tilde{\Delta}^2(j_1, \ldots, j_n) = \frac{\Delta^2(x_{j_1}, \ldots, x_{j_n})}{\sum\limits_{j_1=1}^{N} \cdots \sum\limits_{j_n=1}^{N} \Delta^2(x_{j_1}, \ldots, x_{j_n})}$$

durchgeführt. Wählt man der Einfachheit halber $q = 1$, so ergibt sich nämlich

$$a_1 = \frac{\sum \cdots \sum \Delta(\zeta; \boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_n})\, \Delta(\boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_n})}{\sum \cdots \sum \Delta^2(\boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_n})}.$$

Weil nun

$$\frac{\Delta(\zeta; \boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_n})}{\Delta(\boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_n})} = a_1(j_1, \ldots, j_n)$$

gilt, ist damit alles gezeigt.

Ein analoges Resultat ist auch gültig, wenn die $p_j \neq 1$ sind. Daraus folgt, daß die Sätze 5.3 und 5.4 ein unendlichdimensionales Analogon zur Methode der kleinsten Quadrate beschreiben. Möglich ist auch eine „zweifache Mittelung" oder die Bildung eines Zufallsmodells für die Methode der kleinsten Quadrate, wenn die Punkte $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ zufällig gewählt werden. Um diesen Fall zu betrachten, beweisen wir zunächst ein einfaches Lemma.

Lemma 5.1. *Es gilt die Identität*

$$\int \mu_N(dQ) \det \|[\varphi_k, \psi_l]\| = \frac{N!}{(N-n)!} \det \|(\varphi_q, \psi_r)\| \tag{5.3.9}$$

mit $Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$.

Der Beweis des formulierten Lemmas folgt aus dem Lemma 4.1, wenn man anstelle von $\det \|[\varphi_k, \psi_l]\|$ den Ausdruck (5.3.6) einsetzt und die Integration ausführt. Dabei sind die Integrale über jeden Summanden der sich ergebenden Summe untereinander gleich und ihre Anzahl beträgt $\binom{N}{n}$. Das ergibt die in (5.3.9) benötigte Konstante.

Aus Lemma 5.1 folgt nun sofort

Satz 5.5. *Wenn die* a_q *aus der Bedingung* (5.3.7) *bestimmt werden, wobei die* φ_k, $k = 1, \ldots, n$, *orthonormierte Funktionen sind, die* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ *mit der Dichte*

$(N-n)!\,(N!)^{-1} \det \|[\varphi_k, \psi_l]\|$ *in bezug auf das Maß* $\mu_N(dQ)$ *verteilt und die* ζ_i *unabhängige Realisierungen der Funktion* ζ *sind, dann gilt*

$$\text{A.} \qquad \mathsf{M}_{\omega, Q} a_q = (f, \varphi_q)\,, \qquad q = 1, \ldots, n\,; \tag{5.3.10}$$

$$\text{B.} \qquad \mathsf{D}_{\omega, Q} a_q \leqq \int \left[f(\boldsymbol{x}) - \sum_{i=1}^{n} (f, \varphi_i)\, \varphi_i(\boldsymbol{x})\right]^2 \mu(d\boldsymbol{x}) + \int \sigma^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{5.3.11}$$

mit

$$\sigma^2(\boldsymbol{x}) = \int \zeta^2(\boldsymbol{x}, \omega)\, \mathsf{P}(d\omega) - f^2(\boldsymbol{x})\,.$$

$$\text{C.} \qquad \operatorname{cov}(a_q, a_r) = 0\,, \qquad q \neq r\,.$$

Der Beweis der Behauptungen A und C verläuft völlig analog zu den Beweisen der Sätze 4.3 und 4.4, nur anstatt auf Lemma 4.1 muß man jetzt auf Lemma 5.1 Bezug nehmen. Wir gehen noch auf die Behauptung B ein. Das Zeichen „$\leqq$" gilt in der Ungleichung unabhängig davon, ob das System $\{\varphi_k\}$ regulär ist oder nicht. Dieser Umstand hängt mit der schon erwähnten Nichtregularität eines beliebigen Funktionensystems über einem diskreten Maß zusammen. Der formale Beweis für (5.3.11) ist nicht schwierig. Es sei wieder $q = 1$, dann gilt

$$\mathsf{D} a_1 = \frac{(N-n)!}{N!}\, \mathsf{M}_\omega \int \mu_N(dQ)\, \frac{\{\det \|[\zeta, \varphi_i], [\varphi_2, \varphi_i], \ldots, [\varphi_n, \varphi_i]\|\}^2}{\det \|[\varphi_1, \varphi_i], \ldots, [\varphi_n, \varphi_i]\|} - (\varphi_1, f)^2$$

$$= \frac{(N-n)!}{N!} \int \mu_N(dQ)\, \frac{\left\{\sum\limits_{1 \leqq i_1 < \cdots < i_n \leqq N} \Delta(\zeta; x_{i_1}, \ldots, x_{i_n})\, \Delta(x_{i_1}, \ldots, x_{i_n})\right\}^2}{\sum\limits_{1 \leqq i_1 < \cdots < i_n \leqq N} \Delta^2(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_n})} - (\varphi_1, f)^2\,.$$

Durch Anwendung der Schwarzschen Ungleichung auf den Zähler und durch Kürzen erhalten wir

$$\mathsf{D} a_1 \leqq \frac{(N-n)!}{N!} \int \mu_N(dQ) \sum_{1 \leqq i_1 < \cdots < i_n \leqq N} \Delta^2(\zeta; \boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_n}) - (f, \varphi_1)^2$$

$$= (f, f) - \sum_{i=1}^{n} (f, \varphi_i)^2 + \int \sigma^2(\boldsymbol{x})\, \mu(d\boldsymbol{x}), \tag{5.3.12}$$

was aber der Behauptung (5.3.11) entspricht. Für die anderen q-Werte können analoge Berechnungen durchgeführt werden. Es ist noch zu bemerken, daß die Voraussetzung der Orthonormiertheit der φ_i in dem selben Sinn wie in den Sätzen 4.3 und 4.4 abgeschwächt werden kann. So ergibt sich also ein enger Zusammenhang zwischen der Methode der kleinsten Quadrate und der stochastischen Interpolation. Dieser Zusammenhang ist für uns weiterhin von Wichtigkeit, obgleich man dabei einschränken muß, daß die unmittelbare praktische Anwendung der Ergebnisse des Satzes 5.5 zu keiner in irgendeinem Sinn optimalen Prozedur führt.

Schließlich gehen wir noch auf einen Fall ein, der von bedeutendem praktischen Interesse ist, der aber andererseits nicht leicht zu untersuchen ist. Wir setzen voraus, daß die Approximation der Funktion $f(\boldsymbol{x})$ mit Hilfe einer nichtlinearen Funktion $G(\boldsymbol{x}, a_1, \ldots, a_n)$ verwirklicht wird, die als genügend glatt angenommen wird. Zur Vereinfachung der Berechnungen setzen wir $n = 2$.

Nun wird

$$\mathsf{M}_\omega \int [\zeta(\boldsymbol{x}, \omega) - G(\boldsymbol{x}, a_1, a_2)]^2 \mu(d\boldsymbol{x}) = \min \tag{5.3.13}$$

gefordert. Wenn die zweiten Ableitungen der Funktion G nach den Parametern a_1 und a_2 existieren und stetig sind, erhält man notwendige Bedingungen für das Minimum in (5.3.13) in der Form

$$\left.\begin{aligned}\int \mu(d\boldsymbol{x}) \, [(f(\boldsymbol{x}) - G(\boldsymbol{x}, a_1, a_2)] \frac{\partial G}{\partial a_1} (\boldsymbol{x}, a_1, a_2) = 0\,, \\ \int \mu\,(d\boldsymbol{x}) \, [f(\boldsymbol{x}) - G(\boldsymbol{x}, a_1, a_2)] \frac{\partial G}{\partial a_2} (\boldsymbol{x}, a_1, a_2) = 0\end{aligned}\right\} \tag{5.3.14}$$

mit $f(x) = \mathsf{M}_\omega \zeta(\boldsymbol{x}, \omega)$.

Gewöhnlich benutzt man zur Lösung des Systems (5.3.14) eine Modifikation des Newtonschen Verfahrens, die in folgendem besteht. Es ist eine gewisse Anfangsapproximation a_1^0, a_2^0 gegeben. In der Umgebung dieses Wertes wird der Integrand aus der ersten Gleichung von (5.3.14) in eine Reihe entwickelt. Durch Vernachlässigung der Glieder zweiter und höherer Ordnung ergibt sich eine in $a_1 - a_1^0$ lineare Gleichung

$$\begin{aligned}\int \mu(dx) \Bigg\{\overline{\frac{\partial G}{\partial a_1}} (f - \overline{G}) + \left[\overline{\frac{\partial^2 G}{\partial a_1^2}} (f - \overline{G}) + \left(\overline{\frac{\partial G}{\partial a_1}}\right)^2\right] (a_1 - a_1^0) \\ + \left[\overline{\frac{\partial^2 G}{\partial a_1\, \partial a_2}} (f - \overline{G}) + \overline{\frac{\partial G}{\partial a_1}} \cdot \overline{\frac{\partial G}{\partial a_2}}\right] (a_2 - a_2^0)\Bigg\} = 0\,.\end{aligned}$$

Hier bedeutet der Querstrich, daß die entsprechenden Funktionen an den Stellen $a_1 = a_1^0$, $a_2 = a_2^0$ zu berechnen sind. Wir machen nun die zusätzliche Voraussetzung, daß die Größen

$$\int \mu(d\boldsymbol{x}) \, \overline{\frac{\partial^2 G}{\partial a_1^2}} (f - \overline{G}) \quad \text{bzw.} \quad \int \mu(d\boldsymbol{x}) \, \overline{\frac{\partial^2 G}{\partial a_1\, \partial a_2}} (f - \overline{G})$$

klein gegenüber

$$\int \mu(d\boldsymbol{x}) \left(\overline{\frac{\partial G}{\partial a_1}}\right)^2 \quad \text{bzw.} \quad \int \mu(d\boldsymbol{x}) \, \overline{\frac{\partial G}{\partial a_1}} \cdot \overline{\frac{\partial G}{\partial a_2}}$$

sind. Führt man analoge Berechnungen für die zweite Formel von (5.3.14) durch, so erhält man folgende Formeln für die sukzessive Verbesserung der Parameterwerte a_1 und a_2:

$$\left.\begin{aligned} a_1^{(q)} &= a_1^{(q-1)} + \frac{1}{\Delta_{q-1}} \begin{vmatrix} (f - G^{(q-1)}, \psi_1^{(q-1)}) & (\psi_1^{(q-1)}, \psi_2^{(q-1)}) \\ (f - G^{(q-1)}, \psi_2^{(q-1)}) & (\psi_2^{(q-1)}, \psi_2^{(q-1)}) \end{vmatrix}, \\ a_2^{(q)} &= a_2^{(q-1)} + \frac{1}{\Delta_{q-1}} \begin{vmatrix} (\psi_1^{(q-1)}, \psi_1^{(q-1)}) & (f - G^{(q-1)}, \psi_1^{(q-1)}) \\ (\psi_1^{(q-1)}, \psi_2^{(q-1)}) & (f - G^{(q-1)}, \psi_2^{(q-1)}) \end{vmatrix}. \end{aligned}\right\} \tag{5.3.15}$$

Dabei wird

$$\psi_m^{(q-1)} = \frac{\partial G}{\partial a_m} (\boldsymbol{x}, a_1^{(q-1)}, a_2^{(q-1)})\,, \qquad G^{(q-1)} = G(\boldsymbol{x}, a_1^{(q-1)}, a_2^{(q-1)})$$

gesetzt, und Δ_{q-1} ist die Gramsche Determinante der Funktionen $\psi_1^{(q-1)}(\boldsymbol{x})$ und $\psi_2^{(q-1)}(\boldsymbol{x})$. Es wird weiter vorausgesetzt, daß der Prozeß der Verbesserung der Parameterwerte (5.3.15) für $q \to \infty$ konvergiert und $a_m^{(q)}$ dabei gegen die Werte a_m strebt, die der Gleichung (5.3.13) genügen. Dann kann man bei jedem Rechenschritt die Sätze 5.3 und 5.4 für die Schätzung der Werte $a_m^{(q)}$ anwenden. Dabei wird die Schätzung $\hat{a}_m^{(q)}$ als Mittelwert der Größen

$$\tilde{a}_m^{(q)} = \hat{a}_m^{(q-1)} + \frac{(-1)^{m+1}}{\tilde{\Delta}_{q-1}} \begin{vmatrix} \zeta(\boldsymbol{x}_1, \omega_1) - \tilde{G}^{(q-1)}(\boldsymbol{x}_1) & \psi_m^{(q-1)}(\boldsymbol{x}_1) \\ \zeta(\boldsymbol{x}_2, \omega_2) - \tilde{G}^{(q-1)}(\boldsymbol{x}_2) & \psi_m^{(q-1)}(\boldsymbol{x}_2) \end{vmatrix} \tag{5.3.16}$$

für verschiedene Realisierungen von $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ bestimmt. $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ sind mit der Dichte $W^{(q-1)} = C(\tilde{\Delta}_{q-1})^2$ verteilt, C ist eine Normierungskonstante, und es ist

$$\tilde{\Delta}_{q-1} = \det ||\psi_i^{(q-1)}(x_j)||^2 .$$

Praktisch muß man die Größen $\hat{a}_m^{(q-1)}$ in (5.3.16) durch ihre Schätzwerte ersetzen, d. h., man muß bei jedem Schritt eine genügend große Anzahl von Versuchen machen. Die Schätzung des dabei entstehenden Fehlers ist sehr kompliziert; wir beschränken uns auf die Aussage „es sind genügend viele Versuche zu machen". Das genannte Verfahren hat den Vorzug, daß eine Hypothese über die Anpassung an die gesuchte Funktion leicht nachgeprüft werden kann, was von großer Bedeutung ist. Man kann auch andere Verfahren zur Bestimmung guter (im Sinn von (5.3.13)) Werte a_1 und a_2 benutzen und im abschließenden Schritt die beschriebene Bildung des Zufallsmodells für die Zerlegung der Dispersion in eine zufällige und eine nichtzufällige Komponente anwenden, wie das früher beschrieben wurde.

Die stochastische Interpolation ist also eine Methode zur Approximation von Funktionen im Mittel im Unterschied zu der Methode der abhängigen Versuche. Offensichtlich kann man sie auch zur Schätzung der Werte der Ableitung der Funktion in den Fällen benutzen, in denen es aus irgendeinem Grund unbequem ist, die Methode der abhängigen Versuche anzuwenden. Die bewiesenen Sätze enthalten die Information über die Lage der Interpolationsstützstellen, deren vernünftige Auswahl besonders im mehrdimensionalen Fall ein sehr schwieriges Problem ist. Dieses Problem ist mit klassischen Methoden nur in einer kleinen Anzahl von Fällen lösbar. Bei der praktischen Modellierung mit der Monte-Carlo-Methode interessiert man sich aber sehr oft für gewisse Abhängigkeiten. Dabei kann in vielen Fällen eine vernünftige Hypothese über die Art dieser Abhängigkeit aufgestellt werden. Dann ermöglicht die Methode der stochastischen Interpolation die Schätzung der Parameter und mit einem gewissen zusätzlichen Aufwand die Prüfung der Hypothese. Die Algorithmen, die mit der Verwendung der Dichte W_0 verknüpft sind, führen auf die Lösung linearer algebraischer Gleichungssysteme mit der gleichzeitigen Berechnung ihrer Determinanten, also auf recht einfache Verfahren (bei einer vernünftigen Anzahl von Parametern). Freilich ist W_0 keine genügend „gute" Dichte, und es sind weitere Untersuchungen der Dichten, die der Bedingung der Erwartungstreue genügen, erforderlich. Vom theoretischen Standpunkt her ist das Problem der (in bezug auf ein gewisses Kriterium) optimalen Verteilungsfunktionen der Stützstellen von Quadratur- oder Interpolationsformeln eine Aufgabe der unendlichdimensionalen linearen Optimierung. Der Betrachtung dieser Problemstellung ist der folgende Paragraph gewidmet.

§ 4. Einiges über Probleme der linearen Optimierung

Wie bisher gebrauchen wir die Bezeichnung $Q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$, $m \geqq n$, und nehmen an, daß Q eine nach dem Gesetz $W(Q)$ verteilte Zufallsgröße ist. Um die Verteilung W hervorzuheben, verwenden wir auch die Bezeichnung Q_W. Weiterhin setzen wir zur Vereinfachung $\mathsf{M} = \mathsf{M}_Q \mathsf{M}_\omega$. Die Funktionen $f(\boldsymbol{x})$ und $\mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_W]$ sollen einem linearen normierten Funktionenraum $\mathbf{F}$ angehören. Auf dem Träger des Maßes μ sei eine gewisse Funktion $h(\boldsymbol{x}, Q_W)$ gegeben. Wir nehmen nun an, daß entweder $\boldsymbol{x}$ fest ist oder $h(\boldsymbol{x}, Q) = h(Q)$ nicht von $\boldsymbol{x}$ abhängt.

Das Problem, mit dem wir uns jetzt beschäftigen wollen, besteht in folgendem: Es wird eine bestimmte Verteilung $\overline{U}$ (oder alle $\overline{U}$) aus einer gewissen Menge $\mathbf{U}$ von Verteilungen gesucht derart, daß für beliebiges $f \in \mathbf{F}$ gilt:

$$\text{A.} \qquad \|f(\boldsymbol{x}) - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_{\overline{U}}]\|_{\mathbf{F}} = \inf_{U \in \mathbf{U}} \|f(\boldsymbol{x}) - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_U]\|_{\mathbf{F}}\,, \tag{5.4.1}$$

$$\text{B.} \qquad \mathsf{M}_{Q_{\overline{U}}} h(\boldsymbol{x}, Q_{\overline{U}}) = \inf_{U \in \overline{\mathbf{U}}} \mathsf{M} h(\boldsymbol{x}, Q_U)\,. \tag{5.4.2}$$

Mit $\overline{\mathbf{U}}$ wird die Menge von Verteilungen aus $\mathbf{U}$ bezeichnet, für die (5.4.1) erfüllt ist. Es kann vorkommen, daß $\overline{\mathbf{U}}$ leer ist, aber ebenso gibt es eine Reihe von Fällen, für die die formulierte Aufgabe sinnvoll ist. Wie bisher soll $\mathscr{L}[\boldsymbol{x}, \zeta, Q]$ eine Linearkombination $\sum_{i=1}^{n} a_i \varphi_i(\boldsymbol{x})$ sein, wobei die φ_i auf dem Träger $\mathfrak{X}_0$ des Maßes μ linear unabhängig sind. Wir nehmen an, daß

$$\mathscr{L}[\boldsymbol{x}, \zeta, Q] = \sum_{q=1}^{n} l_q(\boldsymbol{x})\, \zeta_q \tag{5.4.3}$$

mit

$$l_q(\boldsymbol{x}) = \sum_{i=1}^{n} a_i^{(q)} \varphi_i(\boldsymbol{x})$$

gilt, wobei die ζ_q unabhängige Realisierungen der Zufallsgröße ζ für $\boldsymbol{x} = \boldsymbol{x}_q$ sind, und die Bedingungen

$$\mathscr{L}[\boldsymbol{x}, \varphi_i(\boldsymbol{x}), Q] = \varphi_i(\boldsymbol{x})\,, \qquad i = 1, 2, \ldots, n\,, \tag{5.4.4}$$

erfüllt sind. Die von uns bereits betrachteten konkreten Funktionen $\mathscr{L}\,[\boldsymbol{x}, \zeta, Q]$ haben offensichtlich die Gestalt (5.4.3), und für sie sind die Bedingungen (5.4.4) erfüllt. Wir machen nun noch einige zusätzliche Voraussetzungen. Der Träger $\mathfrak{X}_0$ des Maßes μ sei kompakt, $f(\boldsymbol{x})$ und die $\varphi_j(\boldsymbol{x})$ gehören dem Raum $\mathbf{C}_{\mathfrak{X}_0}$ der über $\mathfrak{X}_0$ stetigen Funktionen an. Folglich gilt auch $\varphi_j \in \mathbf{L}_2(\mu)$.

Wir setzen in (5.4.1) $\mathbf{F} = \mathbf{L}_2(\mu)$. Wegen $\mathsf{M}_\omega \mathscr{L}[\boldsymbol{x}, \zeta, Q] = \mathscr{L}[\boldsymbol{x}, f, Q]$ gilt

$$\inf_U \|f(\boldsymbol{x}) - \mathsf{M}\mathscr{L}[x, \zeta, Q_U]\|_{\mathbf{L}_2(\mu)} = \inf_U \|f(\boldsymbol{x}) - \mathsf{M}_{Q_U}\mathscr{L}[\boldsymbol{x}, f, Q]\|_{\mathbf{L}_2(\mu)}$$

$$\geqq \|f(\boldsymbol{x}) - \sum_{j=1}^{n} (f, \varphi_j)\, \varphi_j(\boldsymbol{x})\|_{\mathbf{L}_2(\mu)}\,.$$

Daher wird (5.4.1) für alle U erfüllt, für die

$$\mathsf{M}_{Q_U} a_q = (f, \varphi_q) \quad \text{für alle} \quad f \in C_{\mathfrak{X}_0}\,, \qquad q = 1, \ldots, n\,, \tag{5.4.5}$$

gilt, und die Aufgabe ist zurückgeführt auf die Bestimmung aller u_0 mit

$$\sum_{j=1}^{n} \varphi_j(\boldsymbol{x}) \left\{\int a_q u_0(dQ) - \int f(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right\} = 0 \tag{5.4.6}$$

und

$$\int h(\boldsymbol{x}, Q_{u_0})\, u_0(dQ) = \inf_U \int h(\boldsymbol{x}, Q_U)\, U(dQ)\,. \tag{5.4.7}$$

Es sei nun $m = n$ und a_q nach den Formeln (5.3.3) bestimmt (Fall der Interpolation) Eine natürliche Forderung ist dann die absolute Stetigkeit von u_0 bezüglich W_0. Für die genaue Formulierung des zugehörigen linearen Optimierungsproblems und des dazu dualen Problems führen wir folgende Funktionenräume ein:

1. Den Raum $\mathbf{W}$ der Maße, die von beschränkter Variation und bezüglich W_0 absolut stetig sind.

2. Den Raum $\mathbf{C}^n_{\mathfrak{X}_0}$ der Vektorfunktionen $\boldsymbol{f} = \big(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)\big)$ mit der Norm $\max\limits_{\boldsymbol{x}_i \in \mathfrak{X}_0} |f(\boldsymbol{x}_i)|$.

3. Den Raum $\mathbf{F}_{W_0}$ der Funktionen $g(Q)$, so daß $\Delta^2(Q)\, g(Q)$ stetig über $\mathfrak{X}_0$ ist und die Norm $\max\limits_{x_i \in \mathfrak{X}_0} |\Delta^2(Q)\, g(Q)|$ existiert.

4. Den Raum $\mathbf{V}$ der Vektormaße $\boldsymbol{v} = \big(v_1(x_1), \ldots, v_n(x_n)\big)$ von beschränkter Variation, der zu $\mathbf{C}^n_{\mathfrak{X}_0}$ adjungiert ist.

Es ist offensichtlich, daß $\mathbf{F}_{W_0}$ zu $\mathbf{W}$ adjungiert ist. Wir bezeichnen nun mit A den Operator, der jedem $\boldsymbol{f} \in \mathbf{C}^n_{\mathfrak{X}_0}$ die Funktion $\mathcal{L}[\boldsymbol{x}, f, Q] \in \mathbf{F}_{W_0}$ zuordnet, und setzen voraus, daß $h(\boldsymbol{x}, Q) \in \mathbf{F}_{W_0}$ ist. Dann besteht die Aufgabe (5.4.4), (5.4.5) in der Bestimmung der u, die das Funktional $\int h(\boldsymbol{x}, Q_\mu)\, u(dQ) = (h, u)$ unter den Nebenbedingung $(A\boldsymbol{f}, u) = (\boldsymbol{f}, {}^tAu) = (\boldsymbol{f}, \boldsymbol{v}_0)$ zu einem Minimum machen. Dabei ist $\boldsymbol{v}_0$ ein festes Element aus $\mathbf{V}$ und ${}^tA$ der zu A adjungierte (duale) Operator. Die dazu duale Aufgabe besteht in der Bestimmung einer Funktion $\boldsymbol{\pi} = \big(\pi(\boldsymbol{x}_1), \ldots, \pi(\boldsymbol{x}_n)\big) \in \mathbf{C}^n_{\mathfrak{X}_0}$, für die

$$\sum_{j=1}^{n} \varphi_j(\boldsymbol{x}) \int_{\mathfrak{X}_0} \pi(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{5.4.8}$$

unter der Bedingung

$$\mathcal{L}[\boldsymbol{x}, \pi, Q] = \sum_{j=1}^{n} \varphi_j(\boldsymbol{x})\, a_j(f, Q) \leqq h(\boldsymbol{x}, Q) \tag{5.4.9}$$

ein Supremum erreicht. Daher besteht die zu (5.4.6), (5.4.7) duale Aufgabe in der Bestimmung einer Funktion $\boldsymbol{\pi}$, für die die Teilsumme ihrer Fourierreihe unter der Bedingung maximal wird, daß ihr Interpolationspolynom (in bezug auf das W_0-Maß) für fast jede Lage der Stützstellen beschränkt ist. Sowohl die primale als auch die duale Aufgabe haben zulässige Lösungen (im Sinn der linearen Optimierung), d. h. Lösungen, die (5.4.6) bzw. (5.4.9) genügen. Der Satz 4.3 zeigt nämlich, daß W_0 eine zulässige Lösung der primalen Aufgabe ist. Aus Satz 4.5 folgt die Zulässigkeit der Lösungen in einer genügend kleinen Umgebung von W_0. Für die duale Aufgabe kann eine zulässige Lösung leicht unmittelbar angegeben werden, nämlich eine geeignete Linearkombination der $\varphi_j(\boldsymbol{x})$, $j = 1, \ldots, n$. Es ist offensichtlich, daß bei entsprechender Wahl der Konstanten in der Linearkombina-

tion die Ungleichung (5.4.9) erfüllt wird, weil $h(\boldsymbol{x}, Q)$ nach unten beschränkt ist (fast überall dem W_0-Maß nach).

Wir zeigen nun die Existenz von Lösungen der formulierten Aufgabe.

Der Beweis kann durch verschiedene Überlegungen begründet werden. Wir benutzen Satz 7 der Arbeit von FAN [1] mit einigen Änderungen in der Bezeichnung.

Satz. *Es seien* $\mathbf{E}$ *und* $\mathbf{F}$ *zwei lokalkonvexe topologische Vektorräume,* $A : \mathbf{E} \to \mathbf{F}$ *sei eine stetige lineare Abbildung, für die* $A(\mathbf{E})$ *in* $\mathbf{F}$ *abgeschlossen ist.* $\boldsymbol{K}$ *sei ein abgeschlossener konvexer Kegel in* $\mathbf{F}$, $\eta_0 \in \mathbf{F}$ *und* $v_0 \in \mathbf{E}^*$. *Wir setzen nun voraus, daß* $\xi_0 \in \mathbf{E}$ *und* $u_0 \in \mathbf{F}^*$ *mit*

$$A\xi_0 - \eta_0 \in \tau\text{-Int}\, \boldsymbol{K}, \tag{5.4.10}$$

$$u_0 \in \tau\text{-Int}\, \boldsymbol{K}^0, \qquad {}^tAu_0 - v_0 = 0 \tag{5.4.11}$$

existieren. Dann existiert das Maximum von (v_0, ξ) *unter der Bedingung für* ξ

$$A\xi - \eta_0 \in \boldsymbol{K} \tag{5.4.12}$$

und das Minimum von (u, η_0) *unter der Bedingung für* u

$$u \in \boldsymbol{K}^0, \qquad {}^tAu - v_0 = 0, \tag{5.4.13}$$

und beide Optima sind gleich. (Hier bezeichnet $\boldsymbol{K}^0$ *die Polare von* $\boldsymbol{K}$ *und* τ-Int $\boldsymbol{K}^\circ$ *das Innere von* $\boldsymbol{K}^\circ$ *bezüglich der* τ $(\mathbf{F}^*, \mathbf{F})$-*Topologie.)*

In unserem Fall setzen wir $\mathbf{E} = \mathbf{C}^n_{\mathfrak{X}_0}$, $\mathbf{F} = \mathbf{F}_{W_0}$. Wegen der Stetigkeit der Funktionen $\varphi_j(x)$ auf $\mathfrak{X}_0$ ist der lineare Operator A beschränkt und daher stetig. Die Abgeschlossenheit von $A(\mathbf{E})$ in $\mathbf{F}_{W_0}$ ist offensichtlich. Der abgeschlossene konvexe Kegel $\boldsymbol{K}$ in $\mathbf{F}_{W_0}$ wird durch die Ungleichung $\eta_0 - \eta \geqq 0$ mit $\eta_0 = h(\boldsymbol{x}, P)$ bestimmt, und (5.4.12) ergibt sich dann aus (5.4.9). Die bereits früher eingeführten Bezeichnungen v_0 und η_0 entsprechen denen des Satzes. Man zeigt auch leicht, daß die Funktionen u, die die (nichtleere) Menge der zulässigen Lösungen der primalen Aufgabe bilden, zur Polare von $\boldsymbol{K}$ gehören, wenn (5.4.9) erfüllt ist. Daraus folgt:

Satz 5.6 *Sind die* $\varphi_j(\boldsymbol{x})$, $j = 1, \ldots, n$ *auf* $\mathfrak{X}_0$ *stetig und gehört* $h(\boldsymbol{x}, P)$ *als Funktion von* P, $P \in \mathfrak{X}_0$ *zu* $\mathbf{F}_{W_0}$, *dann gibt es ein* $\bar{u} \in \mathbf{F}_{W_0}$, *für das* (5.4.7) *unter der Bedingung* (5.4.6) *erfüllt ist. Weiterhin existiert ein* $\bar{\pi}(P) \in \mathbf{C}^n_{\mathfrak{X}_0}$, *für das das Funktional* (5.4.8) *unter der Bedingung* (5.4.9) *sein Supremum annimmt, und es gilt*

$$\int h(\boldsymbol{x}, P)\, \bar{u}(dP) = \sum_{j=1}^{n} \varphi_j(\boldsymbol{x}) \int_{\mathfrak{X}} \bar{\pi}(\boldsymbol{x})\, \varphi_j(\boldsymbol{x})\, \lambda(d\boldsymbol{x}). \tag{5.4.14}$$

Die Funktion $\bar{u}$, deren Existenz bewiesen ist, heißt *h-optimal*. Die Approximation der Funktion f mit Hilfe der Methode der kleinsten Quadrate unterscheidet sich offensichtlich in keiner Weise von dem betrachteten Fall, und es gilt ein analoges Resultat. Die Existenz einer zulässigen Lösung der primalen Aufgabe wird dabei durch Satz 5.5 gesichert.

Ist ζ eine Hilbertsche Zufallsfunktion, dann wählt man als Objekt für die Minimierung natürlicherweise die Größe

$$\begin{aligned} \mathsf{M}(\zeta - \mathscr{L}[\boldsymbol{x}, \zeta, Q_u])^2 &= \mathsf{M}(\zeta(\boldsymbol{x}) - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u] + \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u] - \mathscr{L}[\boldsymbol{x}, \zeta, Q_u])^2 \\ &= \mathsf{M}(\zeta(\boldsymbol{x}) - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u]) + \mathsf{M}(\mathscr{L}[\boldsymbol{x}, \zeta, Q_u] - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u])^2 . \end{aligned}$$

Wegen der Erwartungstreue (vgl. (5.4.5)) ist der erste Summand unabhängig von dem gewählten u, und die Aufgabe führt auf die Minimierung der Größe

$$\mathsf{D}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u] = \mathsf{M}(\mathscr{L}[\boldsymbol{x}, \zeta, Q_u] - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u])^2 . \tag{5.4.15}$$

Wenn ein gewisses $\zeta = \zeta_0$ gegeben ist, dann kann man

$$h(\boldsymbol{x}, Q_u) = \mathscr{L}[\boldsymbol{x}, \zeta_0, Q_u] - \mathsf{M}\mathscr{L}[\boldsymbol{x}, \zeta, Q_u]$$

setzen. Eine h-optimale Funktion $\overline{u}$ existiert nach Satz 5.6. Ist sie eindeutig, so ist das Interpolationsverfahren $(\mathscr{L}; \overline{u})$ zulässig im Sinne der Definition des vorigen Paragraphen. Das ist klar, da eine beliebige über $\overline{u}$ dominierende Prozedur u auch eine h-optimale Funktion sein muß, aber diese nach Voraussetzung eindeutig sein soll. Wenn mehrere h-optimale Funktionen u existieren, dann bilden sie eine konvexe Menge, und man kann unter ihnen eine gewisse Menge h_1-optimaler auswählen. Ihre Existenz ergibt sich durch sinngemäße Wiederholung des Beweises von Satz 5.6. Dieser Prozeß kann fortgesetzt werden, und wenn er mit der Bestimmung einer einzigen Verteilungsfunktion abschließt, dann kommen wir ebenfalls zu einer zulässigen Prozedur.

Jetzt gehen wir noch auf die Aufgabe der Integralberechnung ein. Die Formulierung der zugehörigen linearen Optimierungsaufgabe unterscheidet sich in diesem Fall nur wenig von der oben dargestellten Aufgabe. Es sei

$$\varkappa[\zeta] \equiv \frac{\Delta(\zeta, Q)}{\Delta(Q)} \approx \int f(\boldsymbol{x})\, \varphi_1(\boldsymbol{x})\, \mu(d\boldsymbol{x}) , \qquad \Delta(\zeta, Q) = \sum_{i=1}^{n} \Delta_i(Q)\, \zeta_i ,$$

eine Interpolationsquadraturformel, die für die Funktionen φ_j, $j = 1, \ldots, n$, exakt gilt. Die Bestimmung einer h-optimalen Verteilungsfunktion geschieht folgendermaßen. Es wird ein u_0 mit

$$\int h(Q)\, u_0(dQ) = \inf_u \int h(Q)\, u(dQ) \tag{5.4.16}$$

gesucht. Die Bedingung der Erwartungstreue lautet

$$\int \sum_{i=1}^{n} \Delta_i(Q)\, f(\boldsymbol{x}_i)\, \frac{u_0(dQ)}{\Delta(Q)} - \int f(\boldsymbol{x})\, \varphi_1(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = 0 . \tag{5.4.17}$$

Die dazu duale Aufgabe besteht in der Bestimmung einer Funktion

$$\overline{\boldsymbol{\pi}}(Q) = \big(\pi(\boldsymbol{x}_1), \ldots, \pi(\boldsymbol{x}_n)\big) \in \mathbf{C}^n_{\mathfrak{X}_0} ,$$

für die das Funktional

$$\int \pi(\boldsymbol{x})\, \varphi_1(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \tag{5.4.18}$$

unter der Bedingung

$$\sum_{i=1}^{n} \frac{\Delta_i(Q)}{\Delta(Q)}\, \pi(\boldsymbol{x}_i) \leqq h(Q) \tag{5.4.19}$$

sein Supremum annimmt. Alle Voraussetzungen, die in Verbindung mit einer Interpolationsaufgabe gemacht werden, sollen erfüllt sein. Die Existenz einer Lösung der formulierten primalen und dualen Aufgabe wird dann völlig analog wie Satz 5.6 bewiesen. Für diese Lösungen gilt eine der Beziehungen (5.4.14) entsprechende Gleichung

$$\int h(Q)\, u_0(dQ) = \int \pi(\boldsymbol{x})\, \varphi_1(\boldsymbol{x})\, \mu(d\boldsymbol{x}) .$$

Wie immer könnte man anstelle von φ_1 eine beliebige der Funktionen φ_j, $j = 2, \ldots, n$, betrachten. Es bereitet auch keine Schwierigkeiten, anstelle der Interpolationsquadraturformeln eine umfangreichere Klasse von Quadraturformeln zu betrachten. Dann muß man allerdings ein geeignetes Maß bestimmen, das an die Stelle des Maßes W_0 tritt. Der Zusammenhang zwischen den h-optimalen Verteilungsfunktionen und den zulässigen Integrationsverfahren wird analog untersucht wie im Fall der Interpolation. Offenbar können die Ergebnisse von Kap. IV, § 4, in diesem Sinn gedeutet werden, wenn als h die Dispersionen $\mathsf{D}\varkappa[f]$ für irgendwelche festen Funktionen f gewählt werden. Die Funktion $h(Q)$ kann mit einer bestimmten Funktionenklasse in Verbindung gebracht werden. Es sei beispielsweise $x \in [0, 1]$, $\varphi_1(x) = 1$ und $f(x)$ auf $[0, 1]$ stetig differenzierbar. Dann (vgl. Kap. IV, § 1) gibt es für das Restglied $R[f]$ der Quadraturformel eine exakte Abschätzung in der Klasse der differenzierbaren Funktionen:

$$R[f] \leqq M \int_0^1 \left| 1 - t - \sum_i \frac{\Delta_i}{\Delta} \vartheta(x_i - t) \right| dt ,$$

und man kann setzen

$$h(Q) = \int_0^1 \left| 1 - t - \sum_i \frac{\Delta_i}{\Delta} \vartheta(x_i - t) \right| dt .$$

Das entsprechende $u_0(Q)$ stellt die optimale Verteilungsfunktion der Stützstellen für die Klasse der differenzierbaren Funktionen dar.

Im Unterschied zum klassischen Fall wird u_0 hier aus einer linearen Aufgabe bestimmt. Nach Meinung des Autors hat übrigens die dargestellte Theorie eine große Perspektive in Verbindung mit der Approximation von Zufallsfunktionen und der Versuchsplanung.

Wir erörtern nun die Frage nach einer näherungsweisen Lösung der formulierten linearen Optimierungsaufgabe. Aufgaben dieser Art und Methoden zu ihrer näherungsweisen Lösung wurden schon in der Literatur betrachtet, und es gibt in dieser Richtung eine Reihe von Resultaten (Werschik und Temelt [1]). Wir wollen aber hier keine Sätze und Beweise bringen, sondern uns auf eine allgemeine Erörterung der Problemstellung beschränken.

1. Ist μ ein diskretes Maß, das in endlich vielen Punkten konzentriert ist, dann werden (5.4.6)/(5.4.7), (5.4.8)/(5.4.9) und (5.4.16)/(5.4.17), (5.4.18)/(5.4.19) endlichdimensionale lineare Optimierungsaufgaben, deren Lösungsmethoden genügend gut ausgearbeitet sind. Es gibt universelle Programme, die eine Lösung dieser Aufgaben selbst bei einer sehr großen Anzahl von Veränderlichen gestattet. Der Fall des diskreten Maßes kann auch als Approximation des stetigen Falles angesehen werden.

2. Eine andere Methode besteht in folgendem: Wie schon beim Beweis von Satz 4.5 bestimmt man gewisse „Ergänzungen" zur Verteilungsfunktion W_0 derart, daß die Bedingung der Erwartungstreue erfüllt ist und sich der Erwartungswert des gewählten Kriteriums $h(\boldsymbol{x}, Q)$ verkleinert.

Etwas ausführlicher formuliert heißt das: Es wird eine gewisse Menge beschränkter in bezug auf die Veränderlichen $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ symmetrischer Funktionen $v_1(Q), \ldots, v_m(Q)$ gewählt und die Funktion

$$W(Q) = W_0(Q) \left[1 + \sum_{i=1}^{n} b_i v_i(Q)\right]$$

betrachtet. Die Konstanten b_i sollen nun so bestimmt werden, daß

a) $W(Q)$ eine Verteilungsdichte bezüglich des Maßes μ_n ist also

$$1 + \sum_{i=1}^{n} b_i v_i(Q) \geqq 0\,, \qquad \sum_{i=1}^{n} b_i \int v_i(Q)\, W_0(Q)\, \mu_n(dQ) = 0\;;$$

b) $\int h(\boldsymbol{x}, Q)\, W(Q)\, \mu_n(dQ) = \min$.

Bei festen $v_i(Q)$ ist das eine Aufgabe der halbunendlichen Optimierung.

3. Schließlich kann man die Bedingung der Erwartungstreue näherungsweise berücksichtigen. Zum Beispiel kann man annehmen, daß f dem endlichdimensionalen Raum aller Linearkombinationen der Funktionen $\psi_j(\boldsymbol{x})$, $j = 1, \ldots, m$, angehört. Wir setzen voraus, daß die $\psi_j(\boldsymbol{x})$ orthonormiert sind, und betrachten $\mathcal{L}[\boldsymbol{x}, f, Q]$. Weil die Werte von f in $\mathcal{L}$ linear eingehen, gehört $\mathcal{L}[\boldsymbol{x}, f, Q]$ wie auch die Funktion Q zu dem endlichdimensionalen Raum mit der Basis $\mathcal{L}[\boldsymbol{x}, \psi_j, Q] = \mathcal{L}_j[Q]$, $j = 1, \ldots, m$. Unter Benutzung von Lemma 4.1 zeigt man dabei leicht, daß die $\mathcal{L}_j[Q]$ paarweise orthogonal mit dem Gewicht $W_0(Q)$ sind. Die Bedingungen für die Erwartungstreue erhalten jetzt die Gestalt

$$\int \mathcal{L}_j[Q]\, W(dQ) = \sum_{i=1}^{n} \varphi_i(\boldsymbol{x})\, (\varphi_i, \psi_j)\,, \qquad j = 1, \ldots, m\,. \tag{5.4.20}$$

Gesucht ist nun ein diskretes Maß W, für das die Gleichungen (5.4.20) gelten. Die Bestimmung der Punkte, in denen das Maß W konzentriert sein muß, damit das Funktional $\int h(\boldsymbol{x}, Q)\, W(dQ)$ sein Minimum annimmt, führt offensichtlich auf die Bestimmung der Extremwerte einer im allgemeinen nichtlinearen Funktion unter den Bedingungen (5.4.20) und der Nichtnegativitätsbedingung für W. Wenn speziell die Bedingungen (5.4.20) fehlen, dann ist das Funktional $\int h(\boldsymbol{x}, Q)\, W(dQ)$ unter der einzigen Bedingung zu minimieren, daß das Maß W in den Extrempunkten nichtnegativ sein muß.

Diese letzte Aufgabe ist auf das engste mit der in der Statistik betrachteten Aufgabe der Versuchsplanung verknüpft, auf die wir in § 5 näher eingehen. Die bisher erhaltenen Resultate haben dagegen mehr eine prinzipielle Bedeutung. Ihre praktische Anwendung bei Berechnungen mit Hilfe der Monte-Carlo-Methode wird eingeschränkt, weil die der Modellierung zugrunde liegenden Verteilungsfunktionen recht kompliziert sind und es schwierig ist, solche Funktionen zu konstruieren. Zwar könnten diese Schwierigkeiten durch eine unzureichende Untersuchung des Problems bedingt sein. Denn offenbar kann man in konkreten Fällen genügend einfache, angenähert optimale Verteilungsfunktionen konstruieren. Werden aber die Realisierungen der Zufallsgröße ζ nicht auf rechnerischem, sondern auf experimentellem Weg erhalten, dann kann ein größerer Umfang bei den vor-

bereitenden Rechnungen und Untersuchungen durchaus gerechtfertigt sein. Das gilt besonders dann, wenn der mit der Durchführung des Experiments verbundene Aufwand recht groß ist.

§ 5. Über Aufgaben der Versuchsplanung

In § 4 wurde ein Weg zur Lösung des Problems der Versuchsplanung skizziert, der sich etwas von den traditionellen Lösungsverfahren für diese Aufgabe unterscheidet (vgl. etwa den Sammelband „Neue Ideen in der Versuchsplanung"). In diesem Paragraphen gehen wir kurz auf einzelne Ergebnisse auf dem Gebiet der Versuchsplanung ein, die unabhängig von der Monte-Carlo-Methode erhalten wurden, und erläutern ihre Bedeutung für die Monte-Carlo-Methode.

Es ist wichtig, noch einmal zu unterstreichen, daß man in der überwiegenden Mehrzahl der Fälle die Monte-Carlo-Methode als Simulation eines gewissen realen Experiments auffassen kann und daß aus diesem Grund die Bearbeitung von experimentellen Daten und die Methoden der Versuchsplanung eine unmittelbare Beziehung zur Monte-Carlo-Methode haben. Wie schon bemerkt, können die Faktorenanalyse und die Varianzanalyse bei der Untersuchung statistischer Modelle realer Erscheinungen von großem Nutzen sein. Es gibt natürlich auch bestimmte Besonderheiten, die für die Monte-Carlo-Methode charakteristisch sind. Ein reales Experiment ist in der Regel mit vielen Einschränkungen verbunden, die in der Natur der Sache liegen. Bei der Anwendung der Monte-Carlo-Methode sind solche Einschränkungen weniger wichtig, obwohl manchmal auch das Umgekehrte gelten kann. So liefert in den Aufgaben des Strahlendurchgangs die Natur eine sehr viel umfangreichere Statistik als die Monte-Carlo-Methode. Die Anzahl der Teilchen, die wirklich an einem Experiment beteiligt sind, ist wesentlich größer als die Anzahl der Teilchen, die bei der Modellierung zu verfolgen sind. Dennoch hat die Monte-Carlo-Methode auch in diesem Fall bestimmte Vorzüge, da sie ohne weiteres eine Änderung des Modells zuläßt (ausführlicher darüber in Kap. VI).

1. Regressionsanalyse. Wir kommen nun zu dem Problem der Planung eines Regressionsexperiments. Methoden, die mit starken Einschränkungen an die Anzahl der Experimente oder mit der Undurchführbarkeit der Experimente für gewisse Werte der Veränderlichen verbunden sind, werden von uns nicht betrachtet. Wir beschränken uns auf die einfachsten Aufgaben, die eine unmittelbare Beziehung zum Gegenstand dieses Buches haben.

Die Veränderliche $\boldsymbol{x}$, deren Werte beliebig genau bestimmt werden können, sei in einem gewissen Gebiet $\mathcal{D}$ gegeben. Als Ergebnis eines Experiments erhält man Realisierungen einer Funktion $\xi(\boldsymbol{x}) = f(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})$, wobei $\varepsilon(\boldsymbol{x})$ der stochastische Fehler ist. Wir setzen voraus, daß die für jedes $\boldsymbol{x}$ bei N Versuchen erhaltenen Werte $\varepsilon(\boldsymbol{x})$ in der Gesamtheit unabhängige Zufallsgrößen sind, die nicht von $\boldsymbol{x}$ abhängen, also $\varepsilon(\boldsymbol{x}) = \varepsilon$, wobei ε mit den Parametern $(0, \sigma^2)$ normalverteilt ist. Wählen wir also gewisse Werte $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, $\boldsymbol{x}_j \in \mathcal{D}$ für $j = 1, \ldots, N$, so ergeben sich

die Werte $\zeta(\boldsymbol{x}_j) = f(\boldsymbol{x}_j) + \varepsilon_j$. Ist die Funktion f in dem Gebiet $\mathcal{D}$ exakt (oder bis auf einen vernachlässigbaren Fehler) in der Form

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} c_i \varphi_i(\boldsymbol{x}), \qquad c_i \text{ Konstanten,} \tag{5.5.1}$$

mit gegebenen Funktionen $\varphi_i(\boldsymbol{x})$ darstellbar, so ist die Aufgabe auf die Schätzung der Parameter c_i, $i = 1, \ldots, n$, zurückgeführt. Offensichtlich kann man die $\varphi_i(x)$ ohne Beschränkung der Allgemeinheit als orthonormiert voraussetzen.

Nun sei $N > n$. Dann benutzen wir für die Schätzung der c_i die Methode der kleinsten Quadrate, der Schätzwert $\hat{c}_i$ für c_i ergibt sich also aus der Forderung

$$\sum_{j=1}^{N} \left[\zeta_j - \sum_{i=1}^{n} \hat{c}_i \varphi_i(\boldsymbol{x}_j)\right]^2 = \min . \tag{5.5.2}$$

Wir bezeichnen mit $\boldsymbol{c}, \hat{\boldsymbol{c}}, \Xi$ jeweils die Vektoren mit den Komponenten c_i, $\hat{c}_i$ und ζ_i, $i = 1, \ldots, n$, und mit Φ die rechteckige Matrix $\Phi = ((\varphi_i(\boldsymbol{x}_j)))$. Dann ergibt sich aus (5.5.2) $\hat{\boldsymbol{c}} = (\Phi^{\mathsf{T}}\Phi)^{-1}\Phi^{\mathsf{T}}\Xi$. Diesen Ausdruck erhält man leicht unmittelbar oder auch aus (5.3.8). Unter den gemachten Voraussetzungen bezüglich ε ist $\hat{\boldsymbol{c}}$ ein normalverteilter Zufallsvektor mit dem Erwartungsvektor $\boldsymbol{c}$ und der Kovarianzmatrix $B = \sigma^2(\Phi^{\mathsf{T}}\Phi)^{-1}$ (Linnik [1]). Daher hängt die Verteilung des Fehlers von $\hat{\boldsymbol{c}}$ von der Wahl der Punkte $\boldsymbol{x}_j$ ab, und man kann die Frage nach ihrer optimalen Wahl stellen. Die Matrix

$$A = N^{-1}\Phi^{\mathsf{T}}\Phi = \left(\left(N^{-1} \sum_{k=1}^{N} \varphi_i(\boldsymbol{x}_k)\, \varphi_j(\boldsymbol{x}_k)\right)\right)$$

heißt Informationsmatrix. Sie ist mit der Matrix B durch Beziehung $B = N\sigma^2 A^{-1}$ verknüpft. Es ist bemerkenswert, daß die in § 4 betrachtete Verteilungsdichte für die Interpolationsstellen, die nach der Methode der kleinsten Quadrate erhalten wurde, bis auf eine Konstante mit der Determinante der Informationsmatrix übereinstimmt. Einem maximalen Wert dieser Determinante entspricht offensichtlich ein minimaler Wert der Determinante der Kovarianzmatrix. Diese letzte Größe heißt auch verallgemeinerte Dispersion, und sie ist gleich dem Volumen des Streuungsellipoids der Parameterschätzung.

Die am häufigsten gebrauchten Kriterien für die optimale Auswahl der Punkte $\boldsymbol{x}_j$ bestehen in folgendem:

1. Man wählt die $\boldsymbol{x}_j$ so, daß die verallgemeinerte Dispersion der Schätzung minimal wird. Eine Menge solcher (nicht notwendig verschiedener) Punkte, die dieser Forderung genügen, heißt *D-optimaler* Plan. Diese Aufgabenstellung entspricht dem allgemeinen Schema in § 4, wenn man dort die Forderung nach der Erwartungstreue wegläßt und $h = \det A^{-1}$ setzt. Das Fehlen von Bedingungen für die Erwartungstreue steht im Einklang mit der Voraussetzung, daß f ein Element des endlichdimensionalen, durch die φ_i aufgespannten Unterraumes ist.

2. Pläne, die die Länge der größten Achse des Streuungsellipsoids (größter Eigenwert der Kovarianzmatrix) oder die Summe der Längen aller Ellipsoidachsen (Spur der Kovarianzmatrix) minimieren, heißen E-optimal bzw. A-optimal.

3. Setzen wir $\hat{f} = \sum_{i=1}^{n} \hat{c}_i \varphi_i(\boldsymbol{x})$, dann gilt für jedes $\boldsymbol{x}$

$$D\hat{f}(x) = \sum_{i=1}^{n} D\hat{c}_i \varphi_i^2(\boldsymbol{x}) + 2 \sum_{i<j} \operatorname{cov}(c_i, c_j)\, \varphi_i(\boldsymbol{x})\, \varphi_j(\boldsymbol{x}) = \sigma^2 N^{-1} \boldsymbol{\varphi}^{\mathsf{T}} A^{-1} \boldsymbol{\varphi}$$

mit $\boldsymbol{\varphi}^{\mathsf{T}} = (\varphi_1(\boldsymbol{x}), \ldots, \varphi_n(\boldsymbol{x}))$. Ist $\mathcal{D}$ das Gebiet, in dem das Experiment durchgeführt wird, also $\boldsymbol{x}_j \in \mathcal{D}$, und wird $S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sup_{x \in \mathcal{D}} D\hat{f}(\boldsymbol{x})$ gesetzt, dann heißt der $S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ minimierende Plan G-optimal. Die aufgezählten Optimalitätskriterien gehören zu den in der mathematischen Statistik meist verbreiteten und untersuchten, obwohl viele Situationen, die bei der Durchführung eines konkreten Experiments auftreten, die Formulierung eines individuellen Optimalitätskriteriums erfordern.

Die Voraussetzung, daß $f(\boldsymbol{x})$ als Linearkombination gegebener Funktionen φ_i, $i = 1, \ldots, n$, darstellbar ist, ist bei weitem nicht immer gerechtfertigt. In § 4 sind wir von allgemeineren Voraussetzungen ausgegangen. Eine Aufgabe, die der allgemeinen Aufgabe von § 3 ähnlich ist, wurde von Box und Draper [1] betrachtet. Als Optimalitätskriterium wählten sie das Integral

$$\int_{\mathcal{D}} \mathsf{M}[\zeta(\boldsymbol{x}) - \mathsf{M}\zeta(\boldsymbol{x})]^2 \, d\boldsymbol{x} \,. \tag{5.5.3}$$

Daraus ergeben sich dann konkrete Empfehlungen für entsprechende Versuchspläne. Die Forderung der Erwartungstreue wird dabei nicht betrachtet. Wenn die Größe

$$|\mathcal{D}|^{-1} \int_{\mathcal{D}} \mathsf{M}[\zeta(\boldsymbol{x}) - \mathsf{M}\zeta(\boldsymbol{x})]^2 \, d\boldsymbol{x} \,,$$

die mittlere Dispersion von $\zeta(\boldsymbol{x})$ bekannt ist, dann ist das Kriterium (5.5.3) gerechtfertigt.

Da die Betrachtung von Problemen der Versuchsplanung nicht unsere Hauptaufgabe ist, beschränken wir uns auf die aufgezählten Optimalitätskriterien. Ausführlichere Informationen sind in der umfangreichen Spezialliteratur enthalten, z. B. Finney [1], Box und Draper [1], Kiefer [1], Schalajewski [1] und im Sammelband „Neue Ideen in der Versuchsplanung". Für uns war es wichtig, daß die nun darzustellenden Ideen der Dispersionsverkleinerung in der Monte-Carlo-Methode auf bekannte Aufgabenstellungen der Versuchsplanung führen, wobei manchmal sogar von etwas allgemeineren Problemen ausgegangen wird.

Im Zusammenhang mit den D-optimalen Plänen wurde von Kiefer [1] der Begriff des stetigen Planes eingeführt. Wie schon gesagt, können bei der Anwendung der Methode der kleinsten Quadrate als Approximationsverfahren im Fall $N > n$ einige Punkte $\boldsymbol{x}_j$ zusammenfallen, was der wiederholten Beobachtung in ein und demselben Punkt entspricht. Wir werden nur die verschiedenen Punkte $\boldsymbol{x}_{j_k}$ aus der Menge aller Punkte $\boldsymbol{x}_j$ betrachten und jeden durch die Anzahl der Beobachtungen m_k oder die Häufigkeit $m_k/N = p_k$ kennzeichnen. So können wir jedem $\boldsymbol{x} \in \mathcal{D}$ einen gewissen Wert $p(\boldsymbol{x})$ (der auch Null sein kann) zuordnen und finden damit einen stetigen Plan, der durch ein bestimmtes Wahrscheinlichkeitsmaß auf $\mathcal{D}$ dargestellt wird. Solche stetigen Pläne wurden beispielsweise in den oben genannten Arbeiten untersucht. Der hier skizzierte Lösungsweg ist

zwar dem in § 4 ähnlich, er unterscheidet sich aber dadurch, daß in § 3 die gemeinsame Verteilung der $\boldsymbol{x}_j$ betrachtet wurde.

Wir erörtern nun die Anwendung der D-optimalen Pläne in der Monte-Carlo-Methode.

Die Werte $\zeta(\boldsymbol{x})$ können Ergebnisse sein, die beim Lösen einer Aufgabe mit der Monte-Carlo-Methode erhalten wurden, und es möge

$$\mathsf{M}\zeta(\boldsymbol{x}) = f(\boldsymbol{x}) = \sum_{i=1}^{n} c_i \varphi_i(\boldsymbol{x}) \tag{5.5.4}$$

gelten. Wir setzen voraus, daß $\zeta(\boldsymbol{x})$ für jedes $\boldsymbol{x}$ durch eine unabhängige Serie von Versuchen geschätzt wird. Wenn genügend viele solcher Versuche durchgeführt werden, dann ist der Fehler $\varepsilon(\boldsymbol{x}) = \zeta(\boldsymbol{x}) - f(\boldsymbol{x})$ normalverteilt, und $\varepsilon(\boldsymbol{x})$ und $\varepsilon(\boldsymbol{x}')$ sind für $\boldsymbol{x} \neq \boldsymbol{x}'$ unabhängig. Die Größe $\mathbf{D}\varepsilon(\boldsymbol{x})$ kann man als unabhängig von $\boldsymbol{x}$ annehmen, wenn bei den Berechnungen in entsprechender Weise die Anzahl der Versuche für jedes $\boldsymbol{x}$ gewählt wird. Das Problem führt so auf die oben formulierte Aufgabe, und die Anwendung eines D-optimalen Planes ist gerechtfertigt.

Speziell seien $x \in [-1, +1]$ und $\varphi_i(x) = P_{i-1}(x)$ die normierten Legendreschen Polynome vom Grad $i - 1$, $i = 1, ..., n$. Dann ist die Aufgabe, einen D-optimalen Plan für $N \neq nk$ bei ganzzahligem $k > 0$ herzuleiten, noch völlig ungelöst (die allgemeinsten Resultate in dieser Richtung erhielt Schalajewski [1]). Vollständig werden wir hier den Fall $N = nk$ untersuchen. Bei der Anwendung der Monte-Carlo-Methode fällt diese Voraussetzung nicht ins Gewicht.

Es gilt nun folgendes Resultat: Der D-optimale Plan ist in n Punkten konzentriert, die Anzahlen der Wiederholungen des Versuchs in jedem Punkt sind untereinander gleich, also gleich k. Die Lage der Punkte $x_1, x_2, ..., x_n$, in denen der D-optimale Plan konzentriert ist, ist folgende: $x_1 = -1$, $x_n = +1$, die übrigen sind die Nullstellen der Ableitung des Legendreschen Polynoms vom Grad $n - 1$.

Gehört also die Funktion $f(x)$ zum endlichdimensionalen Raum der Polynome von Grad $n - 1$, so führt die Methode der kleinsten Quadrate zur Interpolation und für $N = kn$ ist der Plan D-optimal. Die Interpolationsstellen müssen dazu in der oben genannten speziellen Weise gewählt werden. Es gibt ähnliche Ergebnisse, die den Fall betreffen, daß f in der Form

$$f(x) = \sum_{i=1}^{n} r(x)\, P_{i-1}(x)\,, \qquad r(x) > 0\,,$$

darstellbar ist, wobei die $P_{i-1}(x)$ mit dem Gewicht $r^2(x)$ orthogonale Polynome sind. Diese Ergebnisse (wie auch die für $r(x) \equiv 1$) kann man unter der Voraussetzung $N = kn$ aus einer elektrostatischen Interpretation der Nullstellen von Orthogonalpolynomen erhalten (vgl. etwa Szegö [1]).

1. $r^2(x) = (1 - x)^{q+1}\,(1 + x)^{t+1}$, $x \in [-1, +1]$, $q, t > -1$. Der D-optimale Plan ist in den Nullstellen des Jacobischen Polynoms $P_{n+1}^{(q,t)}$ mit den Parametern q und t konzentriert.

2. $r^2(x) = e^{-x}$, $x \in [0, \infty)$. Der optimale Plan ist in den Nullstellen des Polynoms $xL_n^{(1)}(x)$ konzentriert. Dabei ist $L_n^{(q)}(x)$ das Laguerresche Polynom mit dem Parameter q vom Grad n.

3. $r^2(x) = x^{q+1} e^{-x}$, $x \in [0, \infty)$, $q > -1$. Der optimale Plan ist in den Nullstellen des Polynoms $L_{n+1}^{(q)}(x)$ konzentriert.

4. $r^2(x) = e^{-x^2}$, $x \in (-\infty, \infty)$. Der D-optimale Plan ist in den Nullstellen des Hermiteschen Polynoms H_{n+1} vom Grad $n + 1$ konzentriert. Im Fall einer trigonometrischen Folge führt die Forderung der D-Optimalität zu äquidistanten Interpolationsstellen. Der mehrdimensionale Fall ist wie gewöhnlich wesentlich komplizierter, aber einzelne Ergebnisse, die sich auf eine spezielle Gestalt der Funktion f beziehen, kann man in der Literatur finden.

Sehr oft werden wegen der Schwierigkeit des Problems im mehrdimensionalen Fall zusätzliche Voraussetzungen über die Anordnung der Punkte des Planes gemacht. So kann man zusätzlich die Invarianz des Planes gegenüber gewissen Transformationen der Veränderlichen fordern, z. B. gegenüber Drehung des Koordinatensystems (Drehungsinvarianz, vgl. auch Kap. IV, § 4). Dabei wird vorausgesetzt, daß das Gebiet $\mathcal{D}$ eine bestimmte Zentralsymmetrie besitzt und daß der Koordinatenursprung im Zentrum von $\mathcal{D}$ liegt. Diese zusätzlichen Voraussetzungen vereinfachen die Gestalt der Kovarianzmatrix und folglich auch die Maximierung ihrer Determinante.

Ausführlicher gehen wir auf den Fall ein, daß $f(x)$ eine Linearkombination der Legendreschen Polynome ist:

$$f(x) = \sum_{i=1}^{n+1} c_i P_{i-1}(x) = \sum_{i=1}^{n+1} d_i x^{i-1}.$$

Bei der Schätzung der Koeffizienten d_i ergibt sich für $N = n$, daß $|A|$ gleich dem Quadrat der Vandermondeschen Determinante ist. Man sieht leicht, daß außerdem

$$|A| = \prod_{j=1}^{n} \tilde{P}_n'(x_j)$$

gilt, wobei $\tilde{P}_n$ ein Polynom mit den Nullstellen $x_1, \ldots, x_n$ und dem höchsten Koeffizienten 1 ist. Die Größe von max $|A|$ in Abhängigkeit von n ist bekannt (Schur [1])

$$M_n = \max_{x_1, \ldots, x_n} |A| = \frac{2^2 \cdot 3^3 \cdot \ldots \cdot n^n \cdot 2^2 \cdot 3^3 \cdot \ldots \cdot (n-2)^{n-2}}{3^3 \cdot 5^5 \cdot \ldots \cdot (2n-3)^{2n-3}}, \quad M_2 = 4. \tag{5.5.5}$$

Dem D-optimalen Plan entsprechen folgende Werte x_i, $i=1, \ldots, n$ (für $n = 1, \ldots, 6$):

$n = 1$: $x_{1,2} = \pm 1$;

$n = 2$: $x_{1,3} = \pm 1$, $x_2 = 0$;

$n = 3$: $x_{1,4} = \pm 1$, $x_{2,3} = \pm 1/\sqrt{5}$;

$n = 4$: $x_{1,5} = \pm 1$, $x_{2,4} = \pm \sqrt{3/7}$, $x_3 = 0$;

$n = 5$: $x_{1,6} = \pm 1$, $x_{2,3,4,5} = \pm \sqrt{(1 \pm 2\sqrt{1/7})/3}$;

$n = 6$: $x_{1,7} = \pm 1$, $x_{2,3,5,6} = \pm \sqrt{(15 \pm 2\sqrt{15})/35}$, $x_4 = 0$.

Anhand eines einfachen Beispiels wollen wir uns nun noch von dem Vorteil des D-optimalen Planes überzeugen, wenn es also Aussagen über die Gestalt der Funktion $f(x)$ gibt. Es sei beispielsweise $f(x) = ax + b$ und $\mathcal{D} = [-1, +1]$ das Gebiet, in dem das Experiment durchgeführt wird. Wir setzen $N = 10$ und führen im ersten Fall gemäß dem D-optimalen Plan jeweils fünf Messungen von ζ in den Punkten $+1$ und -1 durch. Im zweiten Fall wird je eine Messung in den äquidistant gelegenen Punkten $\pm 1/9$, $\pm 3/9$, $\pm 5/9$, $\pm 7/9$, ± 1 durchgeführt. Man sieht leicht, daß in beiden Fällen $\operatorname{cov}(\hat{a}, \hat{b}) = 0$ ist, dagegen ist im ersten Fall $\mathbf{D}\hat{a} = \mathbf{D}\hat{b} = \sigma^2/10$ und im zweiten Fall $\mathbf{D}\hat{a} = 27\sigma^2/110 \approx 0{,}245\sigma^2$ und $\mathbf{D}\hat{b} = \sigma^2/10$. Bei Anwendung der Monte-Carlo-Methode im zweiten Fall muß man also die Anzahl der Versuche um das 1,7-fache vergrößern, damit der gleiche maximale Fehler der Funktion $ax + b$ auf $[-1, +1]$ erreicht wird wie beim optimalen Plan. Man kann leicht Beispiele einer noch weniger sinnvollen Lage der Punkte finden, besonders im mehrdimensionalen Fall.

Auf diese Weise können die Methoden der Versuchsplanung ein nützliches Hilfsmittel sein, um die Effektivität der Monte-Carlo-Methode zu erhöhen.

2. Methode der stochastischen Approximation. Diese Methode wurde zuerst in der Arbeit von ROBBINS und MONRO [1] formuliert und hat unmittelbare Beziehung zur Problematik der Versuchsplanung und dem Gegenstand dieses Kapitels. Im Unterschied zu den von uns betrachteten Aufgaben der Versuchsplanung hat das Verfahren der stochastischen Approximation sukzessiven Charakter. Es wird zweckmäßigerweise benutzt, um Kenntnisse lokaler Art über den Erwartungswert einer Zufallsfunktion zu erhalten.

Es sei $\zeta(\boldsymbol{x})$ eine Zufallsfunktion, deren Realisierungen in bestimmten festen Punkten als Ergebnis eines Experiments erhalten werden, und $f(\boldsymbol{x}) = \mathsf{M}\zeta(\boldsymbol{x})$. Wenn man $f(\boldsymbol{x})$ mit Hilfe der bereits dargelegten Methoden approximiert, kann man daraus Aussagen über die Lage der Extremwerte von $f(\boldsymbol{x})$ oder Aussagen über die Lösungen der Gleichung $f(\boldsymbol{x}) = a$ mit gegebenem a erhalten. Ein solches Herangehen ist allerdings in numerischer Hinsicht nicht immer effektiv, und außerdem kann man erwarten, daß die Bestimmung einer lokalen Eigenschaft der Funktion $f(\boldsymbol{x})$ vielleicht weniger Experimente erfordert als die Approximation von $f(\boldsymbol{x})$ im ganzen Gebiet $\mathcal{D}$. Eine Lösung dieser Aufgabe über die lokalen Eigenschaften von $f(\boldsymbol{x})$ kann unter bestimmten Bedingungen mit der Methode der stochastischen Approximation erhalten werden. Darauf gehen wir im weiteren kurz ein.

Jedem Wert x, $-\infty < x < \infty$, entspricht eine Zufallsgröße $\zeta(x)$ mit der Verteilungsfunktion $F(y; x) = \mathcal{P}(\zeta(x) \leq y)$, also $f(x) = \int y \, dF(y; x)$.

Wir wollen nun eine Lösung ϑ der Gleichung $f(x) = a$ durch Meßwerte der Zufallsgröße $\zeta(x_k)$ unter der Bedingung schätzen, daß wir die Größen x_k (die Meßstellen) geeignet auswählen können. Zur Lösung dieser Aufgabe wird eine Folge von Zufallsgrößen x_k, $k = 1, 2, \ldots$, folgenderweise konstruiert:

$$x_{k+1} = x_k + \gamma_k(\zeta(x_k) - a) . \tag{5.5.6}$$

Dabei ist x_1 willkürlich gewählt, und die γ_k bilden eine bestimmte Zahlenfolge mit den Eigenschaften

$$\sum_{k=1}^{\infty} \gamma_k = \infty\,, \quad \sum_{k=1}^{\infty} \gamma_k^2 < \infty\,. \tag{5.5.7}$$

Für genügend großes k kann x_k als Schätzung für ϑ dienen.

Das Verfahren zur Bestimmung der x_k hat also sukzessiven Charakter und ist dem Wesen nach eine Methode der Versuchsplanung mit dem Ziel, ϑ durch Realisierungen der Zufallsgröße $\zeta(x)$ zu bestimmen. Für die γ_k kann man offensichtlich die Folge $\gamma_k = \dfrac{1}{k}$ wählen. Die Bedingungen für die Konvergenz der Folge x_k gegen den Grenzwert ϑ im Mittel und mit der Wahrscheinlichkeit 1 bestehen in folgendem (Blum [1], Dvoretzky [1]):

1. Für alle x und bestimmte feste A und B gilt $|f(x)| \leqq A|x| + B$.
2. $\mathsf{M}(\zeta(x) - f(x))^2 \leqq \sigma^2 < \infty$.
3. $\inf |f(x) - a| > 0$ für alle x mit $1/\varepsilon > |x - \vartheta| > \varepsilon$ und beliebiges ε, $0 < \varepsilon < 1$.

Gladyschew [1] ersetzt die Bedingung 2 durch die Bedingung

2'. $\mathsf{M}(\zeta(x) - f(x))^2 \leqq A(1 + x^2)$ für alle x, die weniger einschränkend als die Bedingung 2 ist.

Ein analoges Verfahren kann man auch im mehrdimensionalen Fall konstruieren. Dazu seien $\zeta_1(x_1, \ldots, x_s), \ldots, \zeta_s(x_1, \ldots, x_s)$ Zufallsgrößen und $f_i(x_1, \ldots, x_s) = \mathsf{M}\zeta_i(x_1, \ldots, x_s)$ die entsprechenden Erwartungswerte. Wir verwenden die Schreibweise $\boldsymbol{x} = (x_1, \ldots, x_s)$, $\boldsymbol{y}(\boldsymbol{x}) = (\zeta_1(\boldsymbol{x}), \ldots, \zeta_s(\boldsymbol{x}))$, $\boldsymbol{a} = (a_1, \ldots, a_s)$ und wollen dann die Lösung $\Theta = (\vartheta_1, \ldots, \vartheta_s)$ des Gleichungssystems

$$f_i(x_1, \ldots, x_s) = a_i\,, \qquad i = 1, \ldots, s\,,$$

schätzen. Das vektorielle Analogon der Gleichung (5.5.6) wird definiert durch das mehrdimensionale Verfahren von Robbins-Monro:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \gamma_k(\boldsymbol{y}(\boldsymbol{x}_k) - \boldsymbol{a})\,, \qquad k = 1, 2, \ldots \tag{5.5.8}$$

Dabei ist $\boldsymbol{x}_1$ ein willkürlicher Vektor und $\{\gamma_k\}$ eine Zahlenfolge, die die Bedingungen (5.5.7) erfüllt.

Die hinreichenden Bedingungen dafür, daß die Norm der Differenz $\boldsymbol{x}_k - \Theta$ im Mittel und mit der Wahrscheinlichkeit 1 gegen Null konvergiert, entsprechen den oben aufgezählten und bilden den Inhalt des folgenden Satzes.

Satz 5.7.

1. *Für alle x und geeignete Konstanten A und B gilt $\|\boldsymbol{f}(\boldsymbol{x})\| \leqq A\|\boldsymbol{x}\| + B$ mit $\boldsymbol{f}(\boldsymbol{x}) = (f_1(x), \ldots, f_s(x))$ und einer Vektornorm $\|\boldsymbol{f}\|$;*

2. *für $1/\varepsilon > \|\boldsymbol{x} - \Theta\| > \varepsilon$, $0 < \varepsilon < 1$, gilt $\inf\,(\boldsymbol{x} - \Theta, \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{a}) > 0$, wobei $(\boldsymbol{a}, \boldsymbol{b})$ das Skalarprodukt der Vektoren $\boldsymbol{a}$ und $\boldsymbol{b}$ bezeichnet;*

3. *es gilt $\mathsf{M}\|\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})\|^2 \leqq \sigma^2 < \infty$.*

Dann gilt

$$\mathsf{P}\Big(\lim_{k\to\infty} \boldsymbol{x}_k = \Theta\Big) = 1 \qquad \textit{und} \qquad \lim_{k\to\infty} \mathsf{M}\|\boldsymbol{x}_k - \Theta\|^2 = 0\,.$$

John, Berkholder und Sacks (siehe Sacks [1]) haben die asymptotische Verteilung des Prozesses der stochastischen Approximation untersucht. Unter Benutzung des Grenzwertsatzes für Martingale bewies Sacks folgende Behauptung (es wird nur der eindimensionale Fall betrachtet).

Satz 5.8. *Es seien folgende Voraussetzungen erfüllt:*

1. $(x - \vartheta)\,(f(x) - a) < 0$;
2. *für alle x und mit geeigneter Konstante A gilt*

$$|f(x) - a| \leqq A|x - \vartheta| \qquad \text{und} \qquad \inf |f(x) - a| > 0$$

mit $t_1 \leqq x - \vartheta \leqq t_2, \qquad 0 < t_1 < t_2 < \infty$;

3. $f(x) = a_0 + a_1(x - \vartheta) + \delta(x, \vartheta) \qquad$ *mit* $\qquad \delta(x, \vartheta) = O(|x - \vartheta)$ *für $x \to \vartheta$ und $a_1 < 0$;*
4. $\sup\limits_x \mathsf{M}(\zeta(x) - f(x))^2 < \infty\,, \qquad \lim\limits_{x \to \vartheta} \mathsf{M}(\zeta(x) - f(x))^2 = s^2 < \infty$;
5. $\lim\limits_{R \to \infty} \; \lim\limits_{\varepsilon \to 0} \; \sup\limits_{|x - \vartheta| < \varepsilon} \int\limits_{|\hat{\varepsilon}(x)| > R} \hat{\varepsilon}^2(x)\, dF < \infty$

mit $\hat{\varepsilon}(x) = \zeta(x) - f(x)$, wobei F die Verteilungsfunktion der Zufallsgröße $\hat{\varepsilon}(x)$ ist;

6. $\gamma_k = \dfrac{A}{k} \quad$ mit $\quad 2Aa_1 > 1$.

Dann gilt mit $\sigma^2 = A^2(2a_1A - 1)^{-1}\, s^2$ für $k \to \infty$

$$\mathsf{P}(k^{1/2}(x_k - \vartheta) \leqq u) \to \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-t^2/2\sigma^2}\, dt\,.$$

Die Bedingung 5 dieses Satzes ist insbesondere dann erfüllt, wenn die Größen $\hat{\varepsilon}(x)$ für verschiedene x unabhängig und identisch verteilt sind oder wenn die Bedingung

$$\sup \mathsf{M}|\hat{\varepsilon}^2(x)|^{2+r} \leqq c < \infty$$

für $|x - \vartheta| < \beta$ mit gewissem $r > 0$, $\beta > 0$ erfüllt ist.

Dvoretzky [1] betrachtet für das Verfahren von Robbins-Monro auch die Aufgabe, eine solche Folge $\{\gamma_k\}$ zu wählen, die die Größe $\mathsf{M}(x_k - \vartheta)^2$ nach einer bestimmten Anzahl von Beobachtungen minimiert. Die von ihm gefundenen Ergebnisse sind Inhalt des folgenden Satzes.

Satz 5.9. *Für die Zufallsgröße $\zeta(x)$, die Funktion $f(x)$ und den Anfangswert x_1 sind folgende Bedingungen erfüllt:*

$$\mathsf{M}(\zeta(x) - f(x))^2 \leqq \sigma^2 < \infty\,,$$
$$0 < m \leqq (f(x) - a)\,(x - \vartheta)^{-1} \leqq M < \infty\,,$$

und

1. $\quad v^2 = \mathsf{M}(x_1 - \vartheta)^2 < 2\sigma^2/m(M - m)\,.$

Dann folgt aus

2. $\quad \gamma_k = mv^2(\sigma^2 + m^2v^2k)^{-1}$

die Aussage

3. $\mathsf{M}(x_{k+1} - \vartheta)^2 \leqq v^2\sigma^2(\sigma^2 + m^2v^2k)^{-1}$.
und die Wahl der Konstanten γ_k ist optimal in dem Sinne, daß bei Nichterfüllung von 2 ein der Bedingung 1 genügender Prozeß existiert, für den die Aussage 3 nicht gilt.

Eine Abschätzung der Konvergenzgeschwindigkeit für die Folge x_k des Verfahrens von ROBBINS-MONRO enthält der folgende Satz.

Satz 5.10. *Sind die Voraussetzungen*

1. $f(x) > a$ *für* $x < \vartheta$, $f(x) < a$ *für* $x > \vartheta$;
2. $c_1|x - \vartheta| \leqq |f(x) - a| \leqq c_2|x - \vartheta|$ *für alle x mit geeignetem c_1 und c_2*;
3. $\mathsf{M}(f(x) - \zeta(x))^2 \leqq s^2 < \infty$;
4. $\gamma_k = Ak^{-1}$ *mit einer Konstanten A*

erfüllt, dann gelten die Abschätzungen

$$\mathsf{M}(x_{k+1} - \vartheta)^2 = \begin{cases} O(k^{-2Ac_2}) & \text{für} \quad 2Ac_2 < 1\,, \\ O(k^{-1}) & \text{für} \quad 2Ac_2 > 1\,, \\ O(k^{-1}\log k) & \text{für} \quad 2Ac_2 = 1\,. \end{cases}$$

Diese Abschätzungen bestätigen, daß unter bestimmten Bedingungen die Benutzung der Methode der stochastischen Approximation in bezug auf den Rechenaufwand günstiger sein kann als die Verfahren, die auf einer Approximation der ganzen Funktion $f(x)$ beruhen.

Ein in vieler Beziehung ähnliches Verfahren wurde von KIEFER und WOLFOWITZ [1] zur Bestimmung der Extremwerte einer Regressionsfunktion vorgeschlagen. Ist $\zeta(\boldsymbol{x}) = \zeta(x_1, \ldots, x_s)$ eine Zufallsgröße, wird $f(\boldsymbol{x}) = \mathsf{M}\zeta(\boldsymbol{x})$ gesetzt und ist $\Theta = (\vartheta_1, \ldots, \vartheta_s)$ ein Extrempunkt von $f(\boldsymbol{x})$, dann wird das Verfahren von KIEFER-WOLFOWITZ zur Bestimmung des Extremwertes durch die Gleichung

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \frac{\beta_k}{\gamma_k}\left(\boldsymbol{y}(\boldsymbol{x}_k, \gamma_k) - \boldsymbol{y}(\boldsymbol{x}_k, -\gamma_k)\right) \tag{5.5.9}$$

definiert. Dabei ist

$$\boldsymbol{x}_k = (x_1^{(k)}, \ldots, x_s^{(k)})\,, \qquad \boldsymbol{y}(\boldsymbol{x}, a) = \left(\zeta(\boldsymbol{x} + a\boldsymbol{e}_1), \ldots, \zeta(\boldsymbol{x} + a\boldsymbol{e}_s)\right),$$

und $\boldsymbol{e}_i$ ist ein Einheitsvektor, dessen i-te Komponente gleich 1 ist. Die Folgen der Konstanten β_k und γ_k müssen den Bedingungen

$$\sum_{k=1}^{\infty} \beta_k = \infty\,, \quad \lim_{k\to\infty} \gamma_k = 0\,, \quad \sum_{k=1}^{\infty} \left(\frac{\beta_k}{\gamma_k}\right)^2 < \infty$$

genügen. Bedingungen für die Konvergenz der Folge (5.5.9) gegen den Extrempunkt Θ werden auch von WILDE [1] angegeben. Wir gehen darauf aber nicht ein. Sowohl das Verfahren von ROBBINS-MONRO wie auch das Verfahren von KIEFER-WOLFOWITZ können nämlich als Spezialfälle eines allgemeineren von DVORETZKY [1] hergeleiteten Verfahrens betrachtet werden. Von ihm wurde folgender Satz bewiesen.

Satz 5.11. *Voraussetzungen:*

1. $\beta_k(r_1, \ldots, r_k)$, $\gamma_k(r_1, \ldots, r_k)$ *seien nichtnegative meßbare Funktionen von $r_1, \ldots, r_k$*;

2. *$\beta_k(r_1, \ldots, r_k)$ gleichmäßig beschränkt und $\lim_{k\to\infty} \beta_k = 0$ gleichmäßig für beliebige r_i;*

3. *$\sum_{k=1}^{\infty} \gamma_k$ gleichmäßig beschränkt und gleichmäßig konvergent in r_i;*

4. *für alle $L > 0$ gibt es nichtnegative Funktionen $\delta_k(r_1, \ldots, r_k)$ mit*

$$\sum_{k=1}^{\infty} \delta_k(r_1, \ldots, r_k) = \infty$$

gleichmäßig für alle r_i mit $\sup |r_i| < L$;

5. *es gibt eine reelle Zahl ϑ und meßbare Abbildungen T_k mit*

$$|T_k(r_1, \ldots, r_k) - \vartheta| \leqq \max\left(\beta_k, (1 + \gamma_k)\,|r_k - \vartheta| - \delta_k\right)$$

für alle reellen r_i;

6. *x_1 und y_k sind Zufallsgrößen mit $\mathsf{M}x_1^2 < \infty$, $\sum_{k=1}^{\infty} \mathsf{M}y_k^2 < \infty$ und $\mathsf{M}(y_k | x_1, \ldots, x_k) = 0$, wobei*

$$x_{k+1} = T_k(x_1, \ldots, x_k) + y_k\,, \qquad k = 1, 2, \ldots, \tag{5.5.10}$$

ist.

Dann gilt

$$\lim_{k\to\infty} \mathsf{M}(x_k - \vartheta)^2 = 0 \qquad \textit{und} \qquad \mathsf{P}\left(\lim_{k\to\infty} x_k = \vartheta\right) = 1\,.$$

Die Ergebnisse dieses Satzes werden in der Arbeit von Kiefer und Wolfowitz [1] bei der Anwendung auf Extremwertaufgaben ausführlich diskutiert.

Die Methode der stochastischen Approximation kann ein wichtiges Hilfsmittel bei der Optimierung der Parameter eines Modells sein. In dieser Beziehung ist eine vor kurzem erschienene Untersuchung von Waisbord und Judin [1] von bedeutendem Interesse, in der ein Markoffscher Prozeß zur Ermittlung des globalen Extremums einer Funktion hergeleitet wird. Ein interessantes Beispiel für die Verbindung der Monte-Carlo-Methode mit der Methode der stochastischen Approximation wird von Botschek [1] angegeben. Er benutzt diese Methoden im Zusammenhang mit Variationsmethoden zur Lösung von Funktionalgleichungen. Auch aus diesem Beispiel kann man auf gewisse Vorzüge der Methode der stochastischen Approximation schließen.

Die Funktionen $f(\boldsymbol{x})$ und $\varphi_1(\boldsymbol{x}), \ldots, \varphi_n(\boldsymbol{x})$ seien Elemente des $\mathbf{L}_2(\mu)$ mit einem Wahrscheinlichkeitsmaß μ. Wir betrachten die Gleichung

$$Lu = f\,, \tag{5.5.11}$$

in der L ein linearer Operator ist und die φ_i zu seinem Definitionsbereich gehören. Es wird vorausgesetzt, daß (5.5.11) eine eindeutige Lösung besitzt. Wir wollen eine Näherungslösung dieser Gleichung in der Gestalt $u_n = \sum_{i=1}^{n} c_i \varphi_i(\boldsymbol{x})$ suchen. Die c_i können aus einer Minimumbedingung für ein gewisses Funktional bestimmt werden. Ein solches Funktional kann beispielsweise $\int [Lu_n - f]^2 \, \mu(d\boldsymbol{x})$ sein (Methode der kleinsten Quadrate). Setzt man $\psi_i(\boldsymbol{x}) = L\varphi_i$, so erhält man zur Bestimmung der c_i ein lineares algebraisches Gleichungssystem

$$\sum_{j=1}^{n} (\psi_i, \psi_j)\, c_j = (\psi_i, f)\,, \qquad i = 1, \ldots, n\,. \tag{5.5.12}$$

Wir führen nun die Bezeichnungen $\boldsymbol{c} = (c_1, \ldots, c_n)$ und $\boldsymbol{\Psi} = (\psi_1(x), \ldots, \psi_n(x))$ ein. Die Skalarprodukte (ψ_i, ψ_j) und (ψ_i, f) bedeuten dann Erwartungswerte der Zufallsgrößen $\psi_i(\omega)\,\psi_j(\omega)$ und $\psi_i(\omega)\,f(\omega)$, wobei die Verteilung der Zufallsgröße ω durch das Maß μ bestimmt wird. Entsprechend der Methode der stochastischen Approximation für die Bestimmung der Lösungen des Gleichungssystems (5.5.8) konvergiert die Folge von Zufallsvektoren

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k + \frac{h}{k}\left[f(\omega_k) - \sum_{j=1}^{n} c_j^{(k)}\psi_j(\omega_k)\right]\boldsymbol{\psi}(\omega_k) \tag{5.5.13}$$

im Mittel und mit der Wahrscheinlichkeit 1 gegen eine Lösung des Systems (5.5.12). Dabei sind die ω_k unabhängige Realisierungen von ω. Ist $\lambda h > 1$ (λ ist der betragskleinste Eigenwert der Matrix des Systems (5.5.12)), dann ergibt sich die Konvergenzordnung $1/k$. Das Rechenverfahren ist sehr einfach und erfordert nur geringen Speicherbedarf der EDVA. (Es ist nicht notwendig, die Matrix des Systems zu speichern.)

Anstelle der Produkte $\psi_i(\omega)\,\psi_j(\omega)$ und $\psi_i(\omega)\,f(\omega)$ kann offensichtlich jede kompliziertere erwartungstreue Schätzung der Größen (ψ_i, ψ_j) und (ψ_i, f) genommen werden. Dieser Umstand macht es möglich, eine große Anzahl zufälliger Komponenten des Iterationsverfahrens (5.5.13) wegzulassen. Es besteht kein prinzipielles Hindernis, analoge Verfahren zur Lösung einer umfangreichen Klasse nichtlinearer Gleichungen mit Hilfe von Variationsmethoden zu benutzen. Dem Autor sind allerdings keine experimentellen Erfahrungen bekannt, auf deren Grundlage er bestimmte Empfehlungen in dieser Richtung geben könnte.

VI. MARKOFFSCHE KETTEN UND DAMIT ZUSAMMENHÄNGENDE PROBLEME

§ 1. Markoffsche Ketten mit endlich vielen Zuständen

Die Modellierung einer Kette mit endlich vielen Zuständen führt dem Wesen nach auf die Modellierung diskreter Zufallsgrößen. Das Modellierungsverfahren hat folgende Form: $\{\boldsymbol{\pi}, \mathcal{P}\}$ bezeichnet eine homogene Markoffsche Kette mit n Zuständen; dabei ist $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)$ die Wahrscheinlichkeitsverteilung des Anfangszustandes und $\mathcal{P} = \left((p_{i,j})\right)$ die Übergangsmatrix (*stochastische Matrix*). Für die Elemente der Matrix $\mathcal{P}$ gilt die Beziehung

$$\sum_{j=1}^{n} p_{i,j} = 1 , \qquad i = 1, ..., n , \qquad p_{i,j} \geqq 0 ,$$

weil jede Zeile mit der Nummer i für sich eine Wahrscheinlichkeitsverteilung darstellt. Ein Objekt, das sich in n Zuständen befinden kann und das durch eine Markoffsche Kette beschrieben wird, wird gewöhnlich zu diskreten Zeitpunkten betrachtet. Mit der Modellierung der Verteilung $\boldsymbol{\pi}$ erhält man den Zustand im Anfangsmoment. Wenn sich danach das Objekt in einem Zustand mit dem Index i befindet, so wird die i-te Zeile der Matrix $\mathcal{P}$ gewählt und die Verteilung modelliert, die dieser Zeile entspricht. Auf diese Weise erhält man eine Folge von Zuständen, die eine Trajektorie der gegebenen Markoffschen Kette bilden.

Die Verallgemeinerung dieses Modellierungsverfahrens auf den Fall einer nichthomogenen Markoffschen Kette, wenn also die Elemente der Matrix $\mathcal{P}$ von der Zeit abhängen, bereitet keine Schwierigkeiten, obwohl die unmittelbare Realisierung eines mit einer nichthomogenen Kette verbundenen Modells auf einer EDVA für großes n manchmal schwierig sein kann.

Die Modellierung einer Markoffschen Kette ist in natürlicher Weise mit der Lösung einer großen Zahl angewandter Aufgaben verbunden. Anschließend gehen wir auf einige davon näher ein.

1. Lösung eines linearen algebraischen Gleichungssystems. Bei der Darstellung eines Lösungsverfahrens für ein lineares algebraisches Gleichungssystem mit der Monte-Carlo-Methode folgen wir im wesentlichen dem zuerst von J. von Neumann und Ulam vorgeschlagenen Weg (Forsythe und Leibler [1]).

Das lineare algebraische Gleichungssystem ist in der Gestalt

$$\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{f} \tag{6.1.1}$$

gegeben. Dabei ist $\boldsymbol{x} = (x_1, ..., x_n)$ der Spaltenvektor der Unbekannten, $\boldsymbol{f} = (f_1, ..., f_n)$ der Vektor der rechten Seiten und $A = \left((a_{i,j})\right)$ die Matrix des Systems.

Wir setzen voraus, daß der betragsgrößte Eigenwert von A kleiner als 1 ist, so daß die Methode der sukzessiven Approximation

$$\boldsymbol{x}^{(k)} = A\boldsymbol{x}^{(k-1)} + \boldsymbol{f} \tag{6.1.2}$$

konvergiert. Wir setzen nun und auch im folgenden $\boldsymbol{x}^{(0)} = \boldsymbol{f}$. Dann gilt

$$\boldsymbol{x}^{(k)} = (A^k + A^{k-1} + \cdots + A + E)\boldsymbol{f}, \tag{6.1.3}$$

und die exakte Lösung $\overline{\boldsymbol{x}}$ des Systems hat die Gestalt

$$\overline{\boldsymbol{x}} = \lim_{k\to\infty} (E + A + \cdots + A^k)\boldsymbol{f} = (E - A)^{-1}\boldsymbol{f}.$$

Wir ordnen nun dem System (6.1.1) eine Markoffsche Kette (nicht unbedingt homogen) mit n Zuständen zu. Die Anfangsverteilung $\boldsymbol{\pi}$ und die Übergangsmatrix $\mathcal{P}_k$ (k Index des Zeitpunktes) müssen gewisse zusätzliche Voraussetzungen erfüllen, die mit dem System (6.1.1) verknüpft sind und die wir im weiteren formulieren. Wir wollen nun das Skalarprodukt $(\boldsymbol{h}, \overline{\boldsymbol{x}})$ mit einem gegebenen Vektor $\boldsymbol{h}$ berechnen. Es ist klar, daß die Berechnung einer Komponente von $\overline{\boldsymbol{x}}$ einen Sonderfall dieser Aufgabe darstellt. Berechnet man gleichzeitig eine gewisse Anzahl von m Skalarprodukten $(\boldsymbol{h}_j, \overline{\boldsymbol{x}})$, $j = 1, \ldots, m$, so erhält man damit bei entsprechender Wahl von m und $\boldsymbol{h}_i$ auch die Lösung $\overline{\boldsymbol{x}}$ des Systems. Wir verbinden nun mit dem System (6.1.1) und dem Vektor $\boldsymbol{h}$ eine bestimmte Markoffsche Kette $\{\boldsymbol{\pi}, \mathcal{P}_k\}$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$, $\mathcal{P}_k = \left((p_{i,j}^{(k)})\right)$, mit

$$\sum_{i=1}^{n} \pi_i = 1, \quad \pi_i \geqq 0, \quad \sum_{j=1}^{n} p_{i,j}^{(k)} = 1, \quad p_{i,j}^{(k)} \geqq 0, \quad i = 1, \ldots, n; \quad k = 1, 2, \ldots$$

für die folgende Bedingungen erfüllt sind:

1. $\pi_i > 0$, falls $h_i > 0$;
2. $p_{i,j}^{(k)} > 0$, falls $a_{i,j}^{(k)} > 0$, $i, j = 1, \ldots, n$.

Wir setzen

$$q_i^{(0)} = \begin{cases} \dfrac{h_i}{\pi_i} & \text{für } \pi_i > 0, \\ 0 & \text{für } \pi_i = 0, \end{cases} \qquad q_{i,j}^{(k)} = \begin{cases} \dfrac{a_{i,j}}{p_{i,j}^{(k)}} & \text{für } p_{i,j}^{(k)} > 0, \\ 0 & \text{für } p_{i,j}^{(k)} = 0. \end{cases} \tag{6.1.4}$$

Nun wird eine natürliche Zahl N vorgegeben und eine Trajektorie der Markoffschen Kette von der Länge N betrachtet. Das durch die Markoffsche Kette beschriebene Objekt werden wir weiterhin oft ein Teilchen nennen und annehmen, daß dieses Teilchen seinen Zustand entsprechend der Trajektorie ändert (sich bewegt). Dem sich bewegenden Teilchen wird ein „Gewicht“ Q_k zugeordnet, das sich während der Bewegung längs der Trajektorie $i_0 \to i_1 \to \cdots \to i_N$ in folgender Weise ändert. Am Anfang, wenn sich das Teilchen im Zustand i_0 befindet, hat es das Gewicht $Q_0 = q_{i_0}^0$, beim Übergang vom Zustand i_0 in den Zustand i_1 wird sein Gewicht gleich $Q_1 = q_{i_0}^0 q_{i_0, i_1}$ usw., allgemein also

$$Q_0 = q_{i_0}^0, \qquad Q_m = Q_{m-1} q_{i_{m-1}, i_m}. \tag{6.1.5}$$

Wir führen nun eine Zufallsgröße ξ_N ein, die auf den Trajektorien der Markoffschen Kette von der Länge N definiert ist:

$$\xi_N = \sum_{m=0}^{N} Q_m f_{i_m}. \tag{6.1.6}$$

Jeder Trajektorie der Länge N entspricht eine Wahrscheinlichkeit $\pi_{i_0} p_{i_0, i_1} \cdots p_{i_{N-1}, i_N}$. Damit ergibt sich der Erwartungswert von ξ_N zu

$$\begin{aligned} \mathsf{M}\xi_N &= \sum_{i_0=1}^{n} \cdots \sum_{i_N=1}^{n} \xi_N \pi_{i_0} p_{i_0, i_1} \cdots p_{i_{N-1}, i_N} \\ &= \sum_{i_0=1}^{n} \cdots \sum_{i_N=1}^{n} \pi_{i_0} p_{i_0, i_1} \cdots p_{i_{N-1}, i_N} \sum_{m=0}^{N} q_{i_0}^0 q_{i_0, i_1} \cdots q_{i_{m-1}, i_m} f_{i_m}. \end{aligned}$$

Setzt man anstelle von $q_{i_0}^0$ und q_{i_{m-1}, i_m} die entsprechenden Werte aus (6.1.4) ein und ändert die Summationsreihenfolge, so ergibt sich

$$\mathsf{M}\xi_N = \sum_{m=0}^{N} \sum_{i_0=1}^{n} \cdots \sum_{i_m=1}^{n} h_{i_0} a_{i_0, i_1} \cdots a_{i_{m-1}, i_m} f_{i_m}.$$

Nach der Definition des Matrizenprodukts gilt

$$A^2 = ((a_{i_0, i_1})) \cdot ((a_{i_1, i_2})) = \left(\left(\sum_{i_1=1}^{n} a_{i_0, i_1} a_{i_1, i_2}\right)\right)$$

und

$$A^m = \left(\left(\sum_{i_1=1}^{n} \cdots \sum_{i_{m-1}=1}^{n} a_{i_0, i_1} \cdots a_{i_{m-1}, i_m}\right)\right).$$

Daraus folgt unmittelbar

$$\mathsf{M}\xi_N = \left(\boldsymbol{h}, \sum_{m=0}^{N} A^m \boldsymbol{f}\right).$$

Für $N \to \infty$ strebt also $\mathsf{M}\xi_N$ gegen $(\boldsymbol{h}, \overline{\boldsymbol{x}})$. Damit haben wir ein mögliches Verfahren der Monte-Carlo-Methode kennengelernt, das mit linearen algebraischen Gleichungssystemen verknüpft ist und das in der Berechnung von Mittelwerten einer Zufallsgröße ξ_N über die Trajektorien der Markoffschen Kette besteht.

Es folgen nun noch einige Bemerkungen, die sich auf das angegebene Verfahren beziehen.

1. Bekanntlich ist $(\boldsymbol{h}, \overline{\boldsymbol{x}}) = (\overline{\boldsymbol{x}}^*, \boldsymbol{f})$, wobei $\overline{\boldsymbol{x}}^*$ die Lösung der Gleichung

$$\boldsymbol{x}^* = A^{\mathsf{T}} \boldsymbol{x}^* + \boldsymbol{h} \tag{6.1.7}$$

mit der zu A transponierten Matrix A^{T} ist. Diese Gleichung kann man übrigens leicht unmittelbar verifizieren. Stellt man $\boldsymbol{h}$ und $\boldsymbol{f}$ durch $\overline{\boldsymbol{x}}^*$ und $\overline{\boldsymbol{x}}$ aus den Gleichungen (6.1.1) bzw. (6.1.7) dar, so ergibt sich $(\overline{\boldsymbol{x}}^* - A^{\mathsf{T}}\overline{\boldsymbol{x}}^*, \overline{\boldsymbol{x}}) = (\overline{\boldsymbol{x}}^*, \overline{\boldsymbol{x}} - A\overline{\boldsymbol{x}})$, was zu der Beziehung $(A^{\mathsf{T}}\overline{\boldsymbol{x}}^*, \overline{\boldsymbol{x}}) = (\overline{\boldsymbol{x}}^*, A\overline{\boldsymbol{x}})$ äquivalent ist. Daraus folgt, daß man bei der Berechnung des Skalarprodukts $(\boldsymbol{h}, \overline{\boldsymbol{x}})$ von der Gleichung $\boldsymbol{x}^* = A^{\mathsf{T}}\boldsymbol{x}^* + \boldsymbol{h}$ ausgehen kann.

Die Markoffschen Ketten, die den beiden Gleichungen (6.1.1) und (6.1.7) entsprechen, können zusammenfallen, wenn für beide gleichzeitig die Bedingungen

(6.1.4) erfüllt sind. Das mit der Gleichung (6.1.7) verknüpfte (adjungierte) Verfahren ist bequemer, wenn man gleichzeitig mehrere Skalarprodukte $(\boldsymbol{h}_j, \overline{\boldsymbol{x}})$, $j = 1, \ldots, m$, berechnen muß. In diesem Fall bleibt die Markoffsche Kette für alle j gleich, und es ändert sich nur die Zufallsgröße ξ_N, so daß die Schätzung ihres Erwartungswertes auf ein und derselben Trajektorie möglich ist.

2. Zur Berechnung der Lösung des Systems (6.1.1) genügt es, n Vektoren für die $\boldsymbol{h}_j$ zu wählen, bei denen die j-te Komponente gleich Eins und die übrigen gleich Null sind.

Bekanntlich kann man die r-te Spalte der Matrix $(E - A)^{-1}$ als Lösung des Systems $\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{h}_r$ erhalten, wenn $\boldsymbol{h}_r$ der r-te Einheitsvektor ist (Spaltenvektor mit r-ter Komponente 1, sonst 0). Daher liefert das mit der Gleichung (6.1.1) verknüpfte Verfahren für $\boldsymbol{f} = \boldsymbol{h}_r$ und $\boldsymbol{h} = \boldsymbol{h}_j$ als Erwartungswert von ξ_N das Element in der r-ten Spalte und j-ten Zeile der Matrix $(E - A)^{-1}$ und das mit der Gleichung (6.1.7) verknüpfte Verfahren das Element in der r-ten Zeile und j-ten Spalte derselben Matrix. Die Differenz $\mathsf{M}\xi_{N+1} - \mathsf{M}\xi_N$ beträgt $(\boldsymbol{h}, A^{N+1}\boldsymbol{f})$ für das erste und $\left(\boldsymbol{f}, (A^\mathsf{T})^{N+1}\boldsymbol{h}\right)$ für das zweite Verfahren. Eine Schätzung dieser Skalarprodukte kann benutzt werden bei der Bestimmung der Eigenwerte der Matrix A, bei der Berechnung der Koeffizienten des charakteristischen Polynoms usw., jeweils entsprechend den bekannten numerischen Verfahren der linearen Algebra (Faddejew und Faddejewa [1]). Die Schätzung der dabei auftretenden stochastischen Fehler erfordert aber zusätzliche Untersuchungen.

3. Aus der Gleichung $\lim_{N\to\infty} \mathsf{M}\xi_N = (\boldsymbol{h}, \boldsymbol{x})$ folgt, daß die Schätzung von ξ_N für eine Trajektorie unendlicher Länge erwartungstreu ist. In der Tat folgt aus der Konvergenz (Konvergenz bezüglich einer gewissen Norm) der Reihe $\sum_{N=0}^{\infty} A^N$, $A^0 = E$, daß von einem gewissen $N = N_0$ ab der Beitrag der bei Vergrößerung von N auftretenden zusätzlichen Summanden zu $\mathsf{M}\xi_{N_0}$ vernachlässigbar klein wird. Praktisch kann die Stelle N_0 der Trajektorie, an der die Berechnung abgebrochen werden sollte, für jede Trajektorie eine andere sein. Man erhält diese Stelle durch eine Schätzung des Beitrages der noch folgenden Glieder in der Reihe (6.1.3). Dabei ist aber eine gewisse Sorgfalt notwendig.

Gilt für wenigstens ein Paar i, j die Beziehung $a_{i,j} = 0$ und damit auch $q_{i,j} = 0$, so wird das Gewicht des Teilchens nach dem Übergang aus den i-ten Zustand in den j-ten Zustand gleich Null. Offensichtlich sind dann auch alle Beiträge bei den nachfolgenden Übergängen gleich Null (das folgt aus (6.1.5)). Ein solcher Übergang hat keinen Sinn. Man kann annehmen, daß das Teilchen beim Übergang vom i-ten Zustand in den j-ten Zustand „absorbiert“ wird und damit seine Trajektorie abbricht. In diesem Fall ist ein künstlicher Abbruch der Trajektorie nicht nötig.

So kann man prinzipiell die Lösung eines linearen algebraischen Gleichungssystems finden, indem man über das durch die Markoffsche Kette bestimmte Maß integriert. Wir kommen damit zu der von uns früher betrachteten Aufgabe der Integralberechnung. Dabei entspricht die Wahl verschiedener Markoffscher Ketten den verschiedenen Varianten der Methode der wesentlichen Stichprobe. Gleichzeitig kann man unter Berücksichtigung der speziellen Struktur des mit der Mar-

koffschen Kette verknüpften Maßes auch gewisse spezielle Resultate über die Dispersion der Schätzung erhalten. Im wesentlichen werden wir diese Resultate unter allgemeineren Bedingungen bei der Untersuchung von Integralgleichungen in § 2 erhalten, und nur eines dieser Resultate, das in unmittelbarer Beziehung zum Lösungsverfahren für lineare algebraische Gleichungssysteme steht, werden wir anschließend anführen.

Zuvor gehen wir noch kurz darauf ein, wann man die Monte-Carlo-Methode zur Lösung von Aufgaben der linearen Algebra anwendet. Praktisch kann die Monte-Carlo-Methode in der beschriebenen Form mit den klassischen Lösungsverfahren unter folgenden Bedingungen konkurrieren:

1. Die Ordnung n der Matrix ist sehr groß, ihre Elemente können einfach berechnet werden, und man will die Lösung des Systems oder die Elemente der Kehrmatrix mit nur geringer Genauigkeit schätzen. Die Größe von n hängt praktisch von den Eigenschaften der Rechenanlage ab, auf der die Aufgabe gelöst wird. Für die heutigen Anlagen muß man $n \approx 10^3$ bis $n \approx 10^4$ als genügend groß ansehen.

2. Die Anwendung der Monte-Carlo-Methode ist besonders günstig, wenn nur eines oder eine kleine Anzahl von Skalarprodukten $(\boldsymbol{h}, \overline{\boldsymbol{x}})$ mit einer geringen Zahl von geltenden Ziffern geschätzt werden soll, da es hierfür nicht nötig ist, vorher die Lösung $\overline{\boldsymbol{x}}$ zu berechnen. Als Beispiel für ein Gleichungssystem, für welches die Anwendung der Monte-Carlo-Methode besonders einfach ist, dient das der Poissonschen Gleichung entsprechende System von Differenzengleichungen. Dieses Beispiel wird in der Literatur über die Monte-Carlo-Methode recht oft betrachtet (z. B. Buslenko u. a. [1], Buslenko und Schreider [1]), und auch wir gehen darauf kurz ein, um dann die oben gestellten Fragen zu klären.

Wir betrachten einen sehr einfachen Fall der ersten Randwertaufgabe für die Gleichung

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y) \tag{6.1.8}$$

im Einheitsquadrat. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß die Randbedingung in der Form $u|_\Gamma = 0$ gegeben ist, wobei Γ der Rand

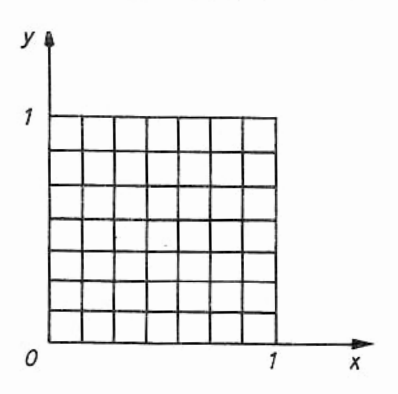


Abb. 6.1

des Quadrates ist. Wir unterteilen das Quadrat durch ein Netz mit der Maschenweite $l = 1/L$ (Abb. 6.1). Wenn man die partiellen Ableitungen in Gleichung (6.1.8) näherungsweise durch die zweiten dividierten Differenzen ersetzt und die Randbedingung berücksichtigt, dann erhält man ein System von $(L-1)^2$ linearen algebraischen Gleichungen der Form

$$u_{i,j} = a_{i,j}u_{i-1,j} + b_{i,j}u_{i,j-1} + c_{i,j}u_{i+1,j} + d_{i,j}u_{i,j+1} - l^2 f_{i,j}\,, \tag{6.1.9}$$

wobei $u_{i,j}$ ein Näherungswert für u im Punkt (il, jl) des Netzes ist, und weiterhin gilt:

$$\left.\begin{aligned} &f_{i,j} = f(il, jl)\,, \quad a_{i,j} = \begin{cases} 1/4 & \text{für} \quad i \neq 1 \\ 0 & \text{für} \quad i = 1\,, \end{cases} \quad b_{i,j} = \begin{cases} 1/4 & \text{für} \quad j \neq 1 \\ 0 & \text{für} \quad j = 1\,, \end{cases} \\ &c_{i,j} = \begin{cases} 1/4 & \text{für} \quad i \neq L-1 \\ 0 & \text{für} \quad i = L-1\,, \end{cases} \quad d_{i,j} = \begin{cases} 1/4 & \text{für} \quad j \neq L-1 \\ 0 & \text{für} \quad j = L-1\,. \end{cases} \end{aligned}\right\} \tag{6.1.10}$$

In den Fällen, in denen die Koeffizienten Null sind, ist es eigentlich gleichgültig zu welchen Unbekannten sie gehören. Man kann annehmen, daß sie zu einer gewissen fiktiven Unbekannten v gehören, die den Randpunkten des Netzes entspricht. Bei einer durchgehenden Numerierung der Unbekannten nach der Regel $u_k = u_{i,j}$ mit $k = j + (L-1)(i-1)$ erhält die Matrix A die in Tabelle 6.1, dargestellte Gestalt. Der Matrix A entspricht in natürlicher Weise die stochastische Matrix von Tabelle 6.2. Die Größen $q_{i,j}$ sind in diesem Fall gleich Eins (außer im Fall des Übergangs in den Zustand 10, wo sie Null sind). Die Konstruktion des Vektors $\boldsymbol{h}$ hängt davon ab, welche Größen wir als Ergebnis der Rechnung erhalten wollen.

Gleichung	Unbekannte									
	1	2	3	4	5	6	7	8	9	10(v)
1		1/4		1/4						
2	1/4		1/4		1/4					
3		1/4				1/4				
4	1/4				1/4		1/4			
5		1/4		1/4		1/4		1/4		
6			1/4		1/4				1/4	
7				1/4				1/4		
8					1/4		1/4		1/4	
9						1 4		1/4		
10 (v)										1

Tabelle 6.1. Gestalt der Matrix A $(L = 4)$

Gleichung	Unbekannte									
	1	2	3	4	5	6	7	8	9	10(v)
1		1/4		1/4						1/2
2	1/4		1/4		1/4					1/4
3		1/4				1/4				1/2
4	1/4				1/4		1/4			1/4
5		1/4		1/4		1/4		1/4		0
6			1/4		1/4				1/4	1/4
7				1/4				1/4		1/2
8					1/4		1/4		1/4	1/4
9						1/4		1/4		1/2
10(v)										1

Tabelle 6.2. Gestalt der Matrix $\mathscr{P}$

Durch Anwendung des adjungierten Verfahrens zur Lösung des Systems (6.6.1) und durch Wahl eines Vektors $\boldsymbol{\pi}$, dessen Komponenten alle gleich $1/(L-1)^2$ sind, gelangen wir zu folgendem Algorithmus. Das Teilchen entsteht gleichwahrscheinlich in einem beliebigen inneren Punkt des Netzes und geht gleichwahrscheinlich in einen beliebigen Nachbarpunkt über. Beim Übergang in einen Randpunkt bricht die Trajektorie ab. Für dieses Lösungsverfahren benötigt man $(L-1)^2$ Speicherplätze. Wenn das Teilchen im k_0-ten Punkt entsteht, wird bei Erreichen des k-ten Punktes (der Anfangspunkt zählt dazu) im k-ten Speicherplatz die Größe $l^2 f_{k_0}$ hinzugefügt. Man verfolgt nun M unabhängige Trajektorien des Teilchens. Der Mittelwert über die Trajektorien vom Inhalt eines jeden Speicherplatzes ist ein Schätzwert für die entsprechende Komponente des Lösungsvektors.

Dieses einfache Verfahren wurde hier absichtlich von einem allgemeineren Standpunkt und daher in etwas komplizierter Weise dargestellt. Der Leser kann jetzt ohne besondere Schwierigkeit Lösungsverfahren für Differenzengleichungen betrachten, die anderen Aufgaben für partielle Differentialgleichungen entsprechen. Im allgemeinen Fall braucht das Teilchen freilich nicht in den Nachbarpunkt des Gitters überzugehen, die Gewichte werden hierbei nach der allgemeinen Formel (6.1.5) berechnet und können ihr Vorzeichen bei einem Übergang ändern.

Aus den Darlegungen folgt, daß bei der Aufstellung eines Lösungsverfahrens mit Hilfe der Monte-Carlo-Methode eine bedeutende Willkür bleibt. Wir gehen insbesondere von der willkürlichen Wahl der Markoffschen Kette $(\boldsymbol{\pi}, \mathscr{P})$ aus. Es ist klar, daß man diese Kette für die Schätzung der uns interessierenden Größen in ungünstiger Weise wählen kann. Wenn uns etwa die k_0-te Komponente x_{k_0} der Lösung interessiert, aber die Wahrscheinlichkeit des Übergangs in den Zu-

stand k_0 sehr klein gewählt wird, dann wird das Teilchen diesen k_0-ten Punkt selten erreichen. Andererseits kann der Beitrag zum Gewicht des Teilchens (das in das Ergebnis eingeht) beim Erreichen des k_0-ten Punktes sehr groß sein. Das hängt davon ab, ob die Dispersion $\mathbf{D}\hat{x}_{k_0}$ der Schätzung von x_{k_0} groß ist. Überlegungen solcher Art können bei der Auswahl der Kette $(\boldsymbol{\pi}, \mathcal{P})$ nützlich sein.

Hat man die Lösung eines linearen algebraischen Gleichungssystems im Auge, so können noch folgende Resultate von Nutzen sein. Wir setzen voraus, daß die Matrix A des Systems

$$\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{f} \tag{6.1.11}$$

symmetrisch ist. Das bedeutet keine wesentliche Einschränkung, da man (wenigstens theoretisch) jedes System mit unsymmetrischer Matrix leicht in ein äquivalentes System mit symmetrischer Matrix überführen kann.

Nach dem direkten Verfahren kann man die Lösung erhalten, indem man dem System (6.1.11) n Markoffsche Ketten $(\boldsymbol{\pi}_j, \mathcal{P})$, $j = 1, \ldots, n$, zuordnet, wobei $\boldsymbol{\pi}_j$ der j-te Einheitsvektor ist und $\mathcal{P}$ nicht von j abhängt. Wie bereits gezeigt wurde, ist es dabei vorteilhaft, bei der Schätzung jeder Unbekannten nicht unabhängige Trajektorien zu benutzen, sondern stets ein und dieselbe. Dazu nimmt man an, daß das Teilchen im i-ten Zustand entstanden ist, falls es während seiner Irrfahrt einmal in diesen Zustand übergeht. Das ist möglich wegen der speziellen Gestalt der Anfangsverteilungen $\boldsymbol{\pi}_j$.

Wenn bei dem üblichen Rechenverfahren $\hat{\xi}_N^{(i)}$ den Schätzwert der i-ten Komponente $\bar{x}_i$ der exakten Lösung über eine Trajektorie der Länge N bezeichnet, dann gilt die Beziehung

$$\hat{\xi}_N^{(i)} = f_i + \frac{a_{i,j}}{p_{i,j}} \hat{\xi}_{N-1}^{(j)}\,, \tag{6.1.12}$$

falls der letzte Schritt der Trajektorie ein Übergang vom j-ten in den i-ten Zustand ist. Wir setzen

$$\hat{\xi}^{(i)} = \lim_{N\to\infty} \hat{\xi}_N^{(i)}\,.$$

Dann gilt das folgende Lemma.

Lemma 6.1. *Ist die Norm der Matrix* $((a_{i,j}^2/p_{i,j}))$ *kleiner als Eins, dann sind die Dispersionen* $\mathbf{D}\hat{\xi}^{(i)}$, $i = 1, \ldots, n$, *endlich und genügen der Gleichung*

$$\mathbf{D}\hat{\xi}^{(i)} = \sum_{j=1}^{n} \frac{a_{i,j}^2}{p_{i,j}} (\mathbf{D}\hat{\xi}^{(j)} + \bar{x}_j^2) - \sum_{j=1}^{n} a_{i,j}\bar{x}_j^2\,. \tag{6.1.13}$$

Beweis. Es gilt

$$\mathbf{D}\hat{\xi}^{(i)} = \mathsf{M}(\hat{\xi}^{(i)})^2 - \bar{x}_i^2 = \lim_{N\to\infty} \mathsf{M}(\hat{\xi}_N^{(i)})^2 - \bar{x}_i^2$$

und wegen (6.1.12)

$$\mathbf{D}\hat{\xi}^{(i)} = \lim_{N\to\infty}\left\{f_i^2 + 2f_i\mathsf{M}\left[\frac{a_{i,j}}{p_{i,j}}\,\hat{\xi}_{N-1}^{(j)}\right] + \mathsf{M}\left[\frac{a_{i,j}}{p_{i,j}}\,\hat{\xi}_{N-1}^{(j)}\right]^2 - \bar{x}_i^2\right\}$$

$$= f_i^2 + 2f_i\sum_{j=1}^{n} a_{i,j}\bar{x}_j + \sum_{j=1}^{n}\frac{a_{i,j}^2}{p_{i,j}}\mathsf{M}(\hat{\xi}^{(j)})^2 - \bar{x}_i^2\,.$$

Die Richtigkeit von (6.1.13) folgt aus dieser letzten Gleichung und aus (6.1.11).

Es ist nun leicht zu sehen, daß mit der formalen Schreibweise

$$p_{i,j} = p_{i,j}^{*} \equiv a_{i,j}\frac{\bar{x}_i}{\sum\limits_{k=1}^{n} a_{i,k}\bar{x}_k} = \frac{a_{i,j}\bar{x}_j}{\bar{x}_i - f_i} \tag{6.1.14}$$

die rechte Seite der Gleichung (6.1.13) verschwindet. Daraus folgt

Satz 6.1. *Ist die durch Gleichung* (6.1.14) *definierte Matrix* $\mathcal{P}^* = ((p_{i,j}^*))$ *eine stochastische Matrix, die die Bedingungen* (6.1.4) (*die* $\mathcal{P}^*$ *mit* A *in Beziehung setzen*) *erfüllt, und die Norm der Matrix* $R = \big((r_{i,j})\big)$ *mit*

$$r_{i,j} = \frac{a_{i,j}\sum\limits_{k=1}^{n} a_{i,k}\bar{x}_k}{\bar{x}_j} = \frac{a_{i,j}(\bar{x}_i - f_i)}{\bar{x}_j} \tag{6.1.15}$$

kleiner als Eins, so gilt

$$\mathbf{D}\hat{\xi}^{(i)} = 0 \qquad \text{für} \qquad i = 1, \ldots, n\,.$$

Gilt weiterhin für alle i

$$\left|1 - \frac{f_i}{\bar{x}_i}\right| < 1\,, \tag{6.1.16}$$

dann ist $\hat{\xi}^{(i)} \equiv \bar{x}_i$ *auf allen Trajektorien der Kette.*

Die Bedingung (6.1.16) *ist offensichtlich erfüllt, wenn alle* f_i *und* $a_{i,j}$ *positiv sind.*

Beweis. Die erste Aussage des Satzes ist offensichtlich richtig. Wir setzen nun $\varepsilon_i = 1 - a_i/\bar{x}_i$. Dann gilt wegen (6.1.6) und (6.1.14) die Beziehung

$$\hat{\xi}_N^{(i)} = \sum_{k=0}^{N}\left\{\prod_{m=1}^{k}\frac{x_{i_{m-1}} - f_{i_{m-1}}}{\bar{x}_{i_m}}\right\} f_{i_k}$$

$$= \bar{x}_i\sum_{k=0}^{N}\left(\prod_{m=1}^{k}\varepsilon_{i_{m-1}}\right)(1 - \varepsilon_{i_m}) = \bar{x}_i\left(1 - \sum_{k=0}^{N}\varepsilon_{i_k}\right).$$

Daraus folgt auch die zweite Aussage des Satzes.

Mit diesem Resultat kann man insbesondere a-priori-Kenntnisse über die Lösung für eine Dispersionsverkleinerung des Schätzwertes $\hat{\xi}^{(i)}$ berücksichtigen. In einer Arbeit [2] benutzt Halton dieses Resultat zur Herleitung eines sukzessiven Lösungsverfahrens für lineare algebraische Gleichungssysteme.

2. Stationäre Verteilungen Markoffscher Ketten. Zunächst ist es zweckmäßig, an einige Definitionen zu erinnern, die sich auf die Struktur einer Markoffschen Kette beziehen.

Man sagt, daß der Zustand γ_k vom Zustand γ_j aus erreichbar ist, (ein echter Nachfolger von γ_j ist), wenn eine Zahl $m > 0$ derart existiert, daß die Wahrscheinlichkeit eines Übergangs von γ_j zu γ_k in m Schritten positiv ist. Eine Menge von Zuständen $\mathscr{C} = (\gamma_{i_1}, \ldots, \gamma_{i_s})$, $s \leqq m$, heißt *abgeschlossen*, wenn von keinem zu $\mathscr{C}$ gehörenden Zustand aus ein nicht zu $\mathscr{C}$ gehörender Zustand erreicht werden kann. Bildet ein einzelner Zustand eine abgeschlossene Menge, so heißt er ein *absorbierender Zustand*. Hat beispielsweise die Übergangsmatrix $\mathscr{P}$ die Gestalt

$$\begin{pmatrix} \mathscr{P}_{1,1} & 0 \\ 0 & \mathscr{P}_{2,2} \end{pmatrix}$$

mit quadratischen stochastischen Matrizen $\mathscr{P}_{1,1}$ und $\mathscr{P}_{2,2}$, dann sind die der Matrix $\mathscr{P}_{1,1}$ entsprechenden Zustände nicht erreichbar von den Zuständen, die $\mathscr{P}_{2,2}$ entsprechen und umgekehrt. Eine Kette mit solchen Matrizen kann offensichtlich nur in dem Fall zur Lösung eines linearen algebraischen Gleichungssystems benutzt werden, in dem das System in zwei unabhängige Gleichungssysteme der entsprechenden Ordnung zerfällt.

Eine Kette heißt *irreduzibel*, wenn es in ihr außer der Menge aller Zustände keine abgeschlossene Menge gibt. Ist die Wahrscheinlichkeit des Übergangs vom j-ten Zustand wieder in den j-ten Zustand (Rückkehr) nach m Schritten für wenigstens ein m ungleich Null, so heißt der größte gemeinsame Teiler t_0 aller m mit dieser Eigenschaft die *Periode* t_0 des j-ten Zustandes. Im Fall $t_0 > 1$ heißt der Zustand selbst periodisch. Besitzt eine Kette periodische Zustände, so heißt sie eine *periodische Kette*.

Es sei nun $i_1 \to i_2 \to \cdots$ eine Trajektorie der Markoffschen Kette. Wir bezeichnen mit $\boldsymbol{p}(N)$ den Vektor mit den Komponenten $\boldsymbol{p}(i_N = 1), \ldots, \boldsymbol{p}(i_N = n)$. Eine für das weitere grundlegende, wichtige Aussage besteht dann in folgendem: Ist eine Markoffsche Kette mit endlich vielen Zuständen irreduzibel und nichtperiodisch, so ist sie *ergodisch*. Das bedeutet, daß unabhängig von der Anfangsverteilung $\boldsymbol{\pi}$ der Grenzwert $\lim\limits_{N\to\infty} \boldsymbol{p}(N) = \overline{\boldsymbol{p}}$ existiert. Die Komponenten des Vektors $\overline{\boldsymbol{p}} = (p_1, \ldots, p_n)$ sind positiv und stellen die stationäre Wahrscheinlichkeitsverteilung dar. Für ergodische Ketten existiert auch der Grenzwert $\lim\limits_{N\to\infty} \mathscr{P}^N = \mathscr{P}_\infty$. Dieser Wert $\mathscr{P}_\infty$ stellt eine stochastische Matrix dar, deren Zeilen alle identisch sind, und $\overline{\boldsymbol{p}}$ genügt der Gleichung $\overline{\boldsymbol{p}} = \overline{\boldsymbol{p}}\mathscr{P}_\infty$. Bei einer unendlichen Anzahl von Zuständen ist auch der Fall $\lim\limits_{N\to\infty} \overline{\boldsymbol{p}}(N) = 0$ möglich. Eine ausführliche Untersuchung dieser Fragen findet der Leser in dem Buch von FELLER [1].

In vielen Aufgaben ist es von Interesse, die stationäre Verteilung einer Markoffschen Kette zu finden oder gewisse mit der stationären Verteilung verbundene Größen zu bestimmen. Grundlegend für die Anwendung der Monte-Carlo-Methode ist bei solchen Aufgaben das folgende Gesetz.

Das starke Gesetz der großen Zahlen für Markoffsche Ketten (SARYMSAKOW [1]): *Ist p die eindeutige stationäre Verteilung der stochastischen Matrix $\mathscr{P}$, so gilt für*

jede Anfangsverteilung $\boldsymbol{\pi}$ *und eine beliebige Funktion der Zustände* $f(\gamma)$ *die Gleichung*

$$\mathsf{P}\left(\lim_{N\to\infty} N^{-1}\sum_{k=1}^{N} f(\gamma_{i_k}) = \sum_{j=1}^{n} \boldsymbol{p}_j f(\gamma_j)\right) = 1\,. \tag{6.1.17}$$

Dabei ist $i_1 \to i_2 \to i_3 \to \cdots$ *die Folge von Übergängen (die Trajektorie) der Markoffschen Kette.*

Für die Schätzung eines Mittelwertes über die stationäre Verteilung $\boldsymbol{p}$ genügt es daher, nur eine Trajektorie der Kette von hinreichend großer Länge zu modellieren.

Die Monte-Carlo-Methode ist ein besonders günstiges Hilfsmittel zur Untersuchung von Markoffschen Ketten mit einer sehr großen Anzahl von Zuständen. Solche Ketten treten beispielsweise in der Bedienungstheorie und in der statistischen Physik auf. Bei der Modellierung ist es in diesen Fällen mitunter nicht leicht, die Matrix $\mathcal{P}$ in expliziter Form anzugeben, besonders wenn die Anzahl der Zustände in die Hunderttausend geht. Jedes genügend komplizierte, provisorische Modell, das in einem Zeitintervall Δt betrachtet wird und das die Eigenschaft einer Markoffschen Kette besitzt, kann bei der Beschreibung der entsprechenden Übergangsmatrix erhebliche Schwierigkeiten bereiten. Das schließt die Möglichkeit aus, die klassischen numerischen Methoden der linearen Algebra zur Berechnung der stationären Verteilung zu verwenden. Was nun die Monte-Carlo-Methode betrifft, so besteht dort die Schwierigkeit in der Schätzung des Verfahrensfehlers.

Wir betrachten jetzt die Aufgabe, die Wahrscheinlichkeit $\mathsf{P}(c)$ einer gewissen Untermenge der stationären Zustände c zu schätzen. Wir erzeugen mit der Monte-Carlo-Methode eine Trajektorie $i_0 \to i_1 \to i_2 \to \cdots$ und nehmen als Schätzwert von $\mathsf{P}(c)$ den Mittelwert

$$\hat{\boldsymbol{P}}(c) = \frac{N_c}{N}, \tag{6.1.18}$$

wobei N_c angibt, wie oft das Teilchen einen zu c gehörenden Zustand eingenommen hat (Fall der Aufgabe C von Kap. I). Die Schätzung (6.1.18) ist erwartungstreu, konsistent und asymptotisch normalverteilt (vgl. Kap. I). Ihre Dispersion kann durch $\mathcal{P}$ und $\mathcal{P}_\infty$ ausgedrückt werden (KEMENY und SNELL [1]):

$$\mathsf{D}\hat{\boldsymbol{P}}(c) = 2B(Z - E)\,Be + o(N^{-2})\,.$$

Dabei ist B eine Diagonalmatrix, auf deren Diagonale die Elemente $b_{ii} = 1$ für $i \in c$ und $b_{ii} = 0$ sonst stehen, $Z = (E - (\mathcal{P} - \mathcal{P}_\infty))^{-1}$ mit der Einheitsmatrix E und $\boldsymbol{e}$ ein Vektor, dessen Komponenten gleich Eins sind. Es ist klar, daß die Anwendung einer solchen Schätzung erst recht in den Fällen unmöglich ist, in denen sogar die explizite Gestalt der Matrix $\mathcal{P}$ nicht zur Verfügung steht. GARABEDJAN und ROMANOWSKI [1] schlugen folgendes Verfahren zur Konstruktion eines Konfidenzintervalls für $\mathsf{P}(c)$ vor. Der Übergang von einem Zustand in einen anderen soll in einem gewissen Zeitintervall erfolgen. Mit $t_1, t_2, \ldots, t_{N_c}$ bezeichnen wir die Momente, in denen das Teilchen in einen Zustand c gelangt, und mit $\tau_i = t_{i+1} - t_i$ die Differenz zwischen diesen Zeiten. Die Folge $\tau_1, \ldots, \tau_{N_c}$ ist stationär, und es gilt $\mathsf{M}\tau_i = \mathsf{M}\tau = 1/\mathsf{P}(c)$. Bildet $[a^-, a^+]$ ein Konfidenzintervall

für $\mathsf{M}\tau$ zu einem gewissen Konfidenzniveau, dann ist offensichtlich $[1/a^+, 1/a^-]$ ein Konfidenzintervall für $\mathsf{P}(c)$ zu demselben Niveau.

Bekanntlich ist die Größe

$$\bar{\tau} = (N_c - 1)^{-1} \sum_{i=1}^{N_c - 1} \tau_i = (t_{N_c} - t_1)(N_c - 1)^{-1} \tag{6.1.19}$$

asymptotisch normalverteilt mit dem Erwartungswert $\mathsf{M}\tau$ und der Dispersion $b/(N_c - 1)$ mit

$$b = \sum_{i=1}^{\infty} \mathsf{M}((\tau_i - \mathsf{M}\tau)(\tau_{i+1} - \mathsf{M}\tau))$$

(Hannan [1], vgl. auch Kap. I). Die Größe b kann über die Folge $\{\tau_i\}$ durch

$$\hat{b}_k = \sum_{s=1}^{k} (N_c - 1 - s)^{-1} \sum_{i=1}^{N_c - 1 - s} (\tau_i - \bar{\tau})(\tau_{i+s} - \bar{\tau})$$

geschätzt werden. Wenn man τ als normalverteilt ansieht, können wir a^+ und a^- bestimmen und ein Konfidenzintervall für $\mathsf{P}(c)$ herleiten.

Die genannte Methode kann man auch benutzen, um den Fehler des Skalarprodukts $(\boldsymbol{h}, \bar{\boldsymbol{p}})$ zu bestimmen, wobei $\boldsymbol{h}$ ein gegebener Vektor mit positiven Komponenten ist. Ausführlicher wird dieses Problem durch Garabedjan und Romanowski [1] untersucht, die auch numerische Beispiele für $n = 20$ angeben.

Es ist klar, daß man bei der Bestimmung der stationären Wahrscheinlichkeiten und der mit ihnen verknüpften Größen genau wie bei der Lösung linearer algebraischer Gleichungssysteme von der Modellierung einer beliebigen ergodischen Markoffschen Kette ausgehen kann. Man erhält dabei die notwendige Wahrscheinlichkeitsverteilung, indem man ein Gewicht des Teilchens nach Formeln einführt, die den Formeln (6.1.4) und (6.1.5) entsprechen.

Prinzipiell mag der Unterschied zwischen den Lösungsmethoden für lineare algebraische Gleichungssysteme und der Bestimmung der stationären Verteilung einer Kette nicht sehr groß erscheinen. Im ersten Fall geht es jedoch um unabhängige Realisierungen der Trajektorien, während es im zweiten Fall um die Berechnung des Mittelwerts abhängiger Zufallsgrößen geht. Freilich ist es auch bei der Berechnung der stationären Verteilung nützlich, wenigstens einige unabhängige Trajektorien zu verfolgen, besonders wenn Zweifel bestehen, ob die Kette irreduzibel und nichtperiodisch ist. Diese letzten Eigenschaften sind bei weitem nicht immer leicht festzustellen. Nützlich kann in dieser Beziehung auch ein von Romanowski [1] beschriebenes Verfahren sein, mit dem man kontrollieren kann, ob die ausgewählten Trajektorien repräsentativ sind.

§ 2. Markoffsche Ketten und Integralgleichungen

Es existiert eine umfangreiche Literatur, die sich mit der Lösung von Integralgleichungen durch die Monte-Carlo-Methode beschäftigt. Eine der ersten Arbeiten, die diesem Problem gewidmet ist, ist die Arbeit von Albert [1] und in der so-

wjetischen Literatur die Arbeit von WLADIMIROW und SOBOL [1]. Eine ausführliche Bibliographie findet sich in den Büchern von BUSLENKO u. a. [1], HAMMERSLEY und HANDSCOMB [1] und auch in der kürzlich erschienenen Monographie von SPANIER und GELBARD [1]. Einige Hinweise auf verhältnismäßig neue Arbeiten werden im weiteren Text gegeben. Vorher ist aber festzustellen, daß die in § 1 beschriebenen Verfahren zur Lösung linearer algebraischer Gleichungssysteme offensichtlich auf den Fall von Integralgleichungen zweiter Art verallgemeinert werden können. Dazu benötigt man aber den Begriff der Markoffschen Kette mit einer beliebigen Anzahl von Zuständen.

Wir betrachten nun eine Integralgleichung der Form

$$\varphi(\boldsymbol{x}) = \int_{\mathcal{D}} K(\boldsymbol{y}, \boldsymbol{x})\, \varphi(\boldsymbol{y})\, d\boldsymbol{y} + f(\boldsymbol{x})\,. \tag{6.2.1}$$

$\mathcal{D}$ ist ein Gebiet des s-dimensionalen euklidischen Raumes, $\boldsymbol{x} \in \mathcal{D}$, $f(\boldsymbol{x})$ und $\varphi(\boldsymbol{x})$ sind auf $\mathcal{D}$ definierte Funktionen, der Kern $K(\boldsymbol{y}, \boldsymbol{x})$ ist auf dem kartesischen Produkt $\mathcal{D} \times \mathcal{D}$ definiert, das Integral in (6.2.1) wird im Lebesgueschen Sinne verstanden. Obwohl auch die Betrachtung etwas komplizierterer Gleichungen (z. B. kann man die Integration allgemeiner definieren, etwa wie in Kap. IV) keine besonderen Schwierigkeiten bereitet, bleiben wir bei dem angegebenen einfachen Fall. An diesem Fall kann man, ohne die Darstellung zu komplizieren, alle prinzipiell wichtigen Fakten erklären.

Für die Gleichung (6.2.1) soll nun die Methode der sukzessiven Approximation konvergieren. d. h., die Lösung $\varphi(\boldsymbol{x})$ ist in Form einer konvergenten Reihe

$$\varphi(\boldsymbol{x}) = \sum_{i=0}^{\infty} \psi_i(\boldsymbol{x}) \tag{6.2.2}$$

mit

$$\psi_0(\boldsymbol{x}) = f(\boldsymbol{x})\,, \qquad \psi_i(\boldsymbol{x}) = \int_{\mathcal{D}} K(\boldsymbol{y}, \boldsymbol{x})\, \psi_{i-1}(\boldsymbol{y})\, d\boldsymbol{y}\,, \qquad i = 1, 2, \ldots$$

darstellbar. Dabei sollen die Funktionen φ und f Elemente eines gewissen Banachraumes $\mathbf{F}$ sein, und die Konvergenz der Reihe (6.2.2) ist in bezug auf die Norm dieses Raumes zu verstehen. Die Norm von Funktionen φ aus $\mathbf{F}$ werden wir wie üblich mit $||\varphi||_{\mathbf{F}}$ oder einfach mit $||\varphi||$ (wenn klar ist, um welchen Raum es sich handelt) bezeichnen. Die Gleichung (6.2.1) läßt sich dann bequem in der Form

$$\varphi = K\varphi + f \tag{6.2.3}$$

schreiben, wenn K den Integraloperator mit dem Kern $K(\boldsymbol{y}, \boldsymbol{x})$ bezeichnet. Wir wollen nun mit der Monte-Carlo-Methode ein Funktional

$$J = \int_{\mathcal{D}} \varphi(\boldsymbol{x})\, h(\boldsymbol{x})\, d\boldsymbol{x} \equiv (\varphi, h)$$

berechnen, wobei $h \in \mathbf{F}^*$ dem zu $\mathbf{F}$ adjungierten Raum (dem Raum der Funktionale) angehört. Das Funktional J kann auch durch die Lösung φ^* der zu (6.2.1) adjungierten Gleichung

$$\varphi^*(\boldsymbol{x}) = \int_{\mathcal{D}} K(\boldsymbol{x}, \boldsymbol{y})\, \varphi^*(\boldsymbol{y})\, d\boldsymbol{y} + h(\boldsymbol{x}) \tag{6.2.4}$$

ausgedrückt werden. Die Gleichung läßt sich analog zu (6.2.3) in der Form

$$\varphi^* = K^*\varphi^* + h$$

schreiben. Genau wie in § 1 zeigt man leicht, daß

$$J = (\varphi, h) = (f, \varphi^*) \tag{6.2.5}$$

gilt, d. h., unsere Aufgabe ist völlig symmetrisch bezüglich der Ausgangsgleichung (6.2.1) und der zu ihr adjungierten Gleichung (6.2.4).

1. Lineare Schätzungen und ihre Erwartungstreue. In der Regel wird den Gleichungen (6.2.1) oder (6.2.4) eine homogene Markoffsche Kette zugeordnet, die durch die Dichten ihrer Anfangsverteilung $\pi(\boldsymbol{x})$ und der Übergangswahrscheinlichkeit $p(\boldsymbol{y} \to \boldsymbol{x})$ gegeben ist. Der Fall einer nichthomogenen Markoffschen Kette ist wenig untersucht, und außer einigen allgemeinen Überlegungen ist es schwer, etwas darüber zu sagen. Die Untersuchung des ebenfalls allgemeineren Falles einer Kette, für die keine Dichte der Anfangsverteilung und der Übergangswahrscheinlichkeit existiert, würde die Darstellung wesentlich komplizieren, ohne viel Neues in prinzipieller Hinsicht zu bringen. Für homogene Ketten $\{\pi, p\}$ muß man einige natürliche Voraussetzungen machen, die mit der Absolutstetigkeit eines später einzuführenden Maßes zusammenhängen:

1. Auf der Menge $\{\boldsymbol{x} : f(\boldsymbol{x}) \neq 0\}$ ist die Funktion $\pi(\boldsymbol{x})$ positiv.

2. Auf der Menge $\{\boldsymbol{x}, \boldsymbol{y} : K(\boldsymbol{y}, \boldsymbol{x}) \neq 0\}$ ist die Funktion $p(\boldsymbol{y} \to \boldsymbol{x})$ positiv. Außerdem ist es günstig,

$$\int_{\mathcal{D}} p(\boldsymbol{y} \to \boldsymbol{x})\, d\boldsymbol{x} = 1 - g(\boldsymbol{y}) , \qquad 0 \leqq g(\boldsymbol{y}) < 1 , \tag{6.2.6}$$

zu setzen und $g(\boldsymbol{y})$ als Wahrscheinlichkeit der Absorption beim Übergang von $\boldsymbol{y}$ nach $\boldsymbol{x}$ aufzufassen.

Entsprechend den allgemeinen Modellierungsmethoden (Kap. II) wird eine Stichprobentrajektorie der Kette folgendermaßen konstruiert. Durch Modellierung einer Zufallsgröße mit der Verteilungsdichte $\pi(\boldsymbol{x})$ findet man den Punkt $\boldsymbol{x}_0$. Dieser Punkt liefert die Anfangskoordinaten des Teilchens, das sich längs der Trajektorie bewegt. Mit einer Wahrscheinlichkeit $g(\boldsymbol{x}_0)$ wird das Teilchen nun absorbiert. Wenn keine Absorption erfolgte, wird die Zufallsgröße $\boldsymbol{x}_1$ mit der Verteilungsdichte $p(\boldsymbol{x}_0 \to \boldsymbol{x}_1)/\big(1 - g(\boldsymbol{x}_0)\big)$ modelliert. Ist also allgemeiner $\boldsymbol{x}_{i-1}$ bekannt, so bricht die Trajektorie mit der Wahrscheinlichkeit $g(\boldsymbol{x}_{i-1})$ ab und wird mit der Wahrscheinlichkeit $1 - g(\boldsymbol{x}_{i-1})$ fortgesetzt. Dann ist $\boldsymbol{x}_i$ eine Realisierung der mit der Dichte $p(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)/\big(1 - g(\boldsymbol{x}_{i-1})\big)$ verteilten Zufallsgröße.

Ist also $g > 0$, dann entspricht der Trajektorie $\boldsymbol{x}_0 \to \boldsymbol{x}_1 \to \cdots \to \boldsymbol{x}_k$ von der Länge k eine Wahrscheinlichkeitsdichte

$$\pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\, g(\boldsymbol{x}_k) .$$

In diesem Fall sind fast alle Trajektorien endlich, und durch die angegebene Dichte auf den endlichen Trajektorien wird ein Maß definiert.

Ist $g \equiv 0$, dann haben fast alle Trajektorien eine unendliche Anzahl von Gliedern. Ein Maß wird auf allen endlichen Abschnitten der Trajektorien $\boldsymbol{x}_0 \to \boldsymbol{x}_1$

$\to \cdots \to \boldsymbol{x}_N$ (auf den „Zylindermengen") definiert. Nach dem Satz von Kolmogoroff kann dieses Maß auf die Menge aller Trajektorien erweitert werden.

Das uns interessierende Funktional J wird nun als Mittelwert einer weitgehend willkürlichen Zufallsgröße über die ausgewählten Trajektorien geschätzt (Schätzung eines Funktionals). In der Regel wird vorausgesetzt, daß diese Schätzung erwartungstreu ist. Wir werden für $g(\boldsymbol{x}) > 0$ eine Schätzung der Gestalt

$$\hat{J}[k] = \sum_{i=0}^{k} Q_i^{(k)}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k)\, h(\boldsymbol{x}_i) = \sum_{i=0}^{k} Q_i^{(k)}(\Xi_k)\, h(\boldsymbol{x}_i) \tag{6.2.7}$$

und für $g(\boldsymbol{x}) \equiv 0$ eine Schätzung der Gestalt

$$\hat{J}[\infty] = \sum_{i=0}^{\infty} Q_i(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots)\, h(\boldsymbol{x}_i) \tag{6.2.8}$$

betrachten. Dabei kennzeichnet k die Stelle, die dem Ende der Trajektorie entspricht; $Q_i^{(k)}$ und Q_i sind gewisse gegebene Funktionen der Trajektorie, die als integrierbar nach dem auf der Menge der Trajektorien definierten Maß vorausgesetzt werden. Schätzungen der Art (6.2.7) und (6.2.8) werden wir linear nennen. Die Existenz ihrer Dispersion wird vorausgesetzt.

Damit die Schätzung erwartungstreu ist, müssen die Funktionen $Q_i^{(k)}$ und Q_i bestimmte Bedingungen erfüllen, die wir jetzt herleiten wollen. Dafür stellen wir das Funktional $J = (\varphi, h)$ in einer für uns bequemen Form dar:

$$(\varphi, h) = \sum_{i=0}^{\infty} \int_{\mathcal{D}} h(\boldsymbol{x})\, \psi_i(\boldsymbol{x})\, dx$$

$$= \int_{\mathcal{D}} h(\boldsymbol{x}) \left\{ f(\boldsymbol{x}) + \sum_{i=1}^{\infty} \int_{\mathcal{D}} d\boldsymbol{x}_1 \cdots \int_{\mathcal{D}} d\boldsymbol{x}_i\, K(\boldsymbol{x}_i, \boldsymbol{x})\, K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i) \cdots K(\boldsymbol{x}_1, \boldsymbol{x}_2)\, f(\boldsymbol{x}_1) \right\} d\boldsymbol{x}\,.$$

Durch eine Änderung der Bezeichnungen erhalten wir die äquivalente Gleichung

$$(\varphi, h) = \sum_{i=0}^{\infty} \int_{\mathcal{R}_i} f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)\, h(\boldsymbol{x}_i)\, d\Xi_i$$

mit $\mathcal{R}_i = \underbrace{\mathcal{D} \times \mathcal{D} \times \cdots \times \mathcal{D}}_{i+1}$ $\mathcal{R}_0 = \mathcal{D}$ und $d\Xi_0 = d\boldsymbol{x}_0$, $d\Xi_i = d\boldsymbol{x}_0\, d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_i$. Der Erwartungswert des Schätzungswertes $\hat{J}[k]$ beträgt dann ($g(\boldsymbol{x})$ sei nicht identisch Null)

$$\mathsf{M}\hat{J}[k] = \sum_{k=0}^{\infty} \int_{\mathcal{R}_k} \hat{J}[k]\, \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\, g(\boldsymbol{x}_k)\, d\Xi_k\,.$$

Die notwendige Bedingung für die Erwartungstreue der Schätzung lautet

$$(\varphi, h) = \mathsf{M}\hat{J}[k]\,. \tag{6.2.9}$$

Unter einigen zusätzlichen Voraussetzungen über die absolute Konvergenz der unendlichen Reihen kann man auch hinreichende Bedingungen für die Erwartungstreue erhalten. Eine der einfachsten Bedingungen dieser Art hat die Gestalt

$$\sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} \Big[f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)\, h(\boldsymbol{x}_k)$$
$$- \sum_{i=0}^{k} Q_i^{(k)}(\Xi_k)\, h(\boldsymbol{x}_i)\, \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\, g(\boldsymbol{x}_k) \Big]\, d\Xi_k = 0\,. \quad (6.2.10)$$

Wir setzen nun

$$Q_i^{(k)} = q_i \omega_i^{(k)} \quad (6.2.11)$$

mit

$$q_0 = \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)}, \qquad q_i = \frac{K(\boldsymbol{x}_0, \boldsymbol{x}_1)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1)} \cdots \frac{K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)}{p(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)}. \quad (6.2.12)$$

Berücksichtigt man $g(\boldsymbol{x}_k) = 1 - \int_{\mathfrak{D}} p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})\, d\boldsymbol{x}_{k+1}$, so kann man die Gleichung (6.2.10) formal in folgender Form schreiben:

$$\int_{\mathfrak{D}} [1 - \omega_0^{(0)}]\, f(\boldsymbol{x}_0)\, h(\boldsymbol{x}_0)\, d\boldsymbol{x}_0$$
$$+ \sum_{k=1}^{\infty} \int_{\mathfrak{R}_k} f(\boldsymbol{x}_0) \Big\{ [1 - \omega_k^{(k)}]\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)\, h(\boldsymbol{x}_k)$$
$$+ \sum_{i=1}^{k-1} [\omega_i^{(k)} - \omega_i^{(k+1)}]\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)\, h(\boldsymbol{x}_i)\, p(\boldsymbol{x}_i \to \boldsymbol{x}_{i+1})$$
$$\cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k) \Big\}\, d\Xi_k = 0\,.$$

Wenn man hier die Integrations- und Summationsreihenfolge vertauscht, ergibt sich folgendes Resultat.

Satz 6.2. *Damit die Schätzung* (6.2.7) *im Fall* $g(\boldsymbol{x}) > 0$ *für* $\boldsymbol{x} \in \mathfrak{D}$ *und bei Erfüllung der Bedingungen*

1. $\sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} |\omega_k^{(k)} f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)\, h(\boldsymbol{x}_k)|\, d\Xi_k < \infty$;
2. $\sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} |(\omega_i^{(k)} - \omega_i^{(k+1)})\, f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1)$
$\cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)\, h(\boldsymbol{x}_i)\, p(\boldsymbol{x}_i \to \boldsymbol{x}_{i-1}) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)|\, d\Xi_k < \infty$

erwartungstreu ist, ist die folgende Gleichung hinreichend:

$$\sum_{i=0}^{\infty} \int_{\mathfrak{R}_i} f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)\, h(\boldsymbol{x}_i) \Big\{ 1 - \omega_i^{(i)} - \sum_{k=i}^{\infty} \int_{\mathfrak{R}_{k+i-1}} [\omega_i^{(k)} - \omega_i^{(k+1)}]$$
$$\times\, p(\boldsymbol{x}_i \to \boldsymbol{x}_{i+1}) \cdots p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})\, d\boldsymbol{x}_{i+1} \cdots d\boldsymbol{x}_{k+1} \Big\}\, d\Xi_i = 0\,. \quad (6.2.13)$$

Ein spezielles Interesse kann der Fall beanspruchen, in dem die $Q_i^{(k)}$ nur von den ersten $i + 1$ Veränderlichen $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_i$ abhängen, da die Rechnung sich dann meistens vereinfacht. Die Gleichung (6.2.13) wird in diesem Fall auch etwas einfacher und kann in folgender Form geschrieben werden:

$$\sum_{i=0}^{\infty} \int_{\mathfrak{R}_i} f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)\, h(\boldsymbol{x}_i) \Bigg\{1 - \omega_i^{(i)} - \sum_{k=i}^{\infty} [\omega_i^{(k+1)} - \omega_i^{(k)}]$$

$$\times \int_{\mathfrak{D}} d\boldsymbol{x}_{i+1} \cdots \int_{\mathfrak{D}} d\boldsymbol{x}_{k+1}\, p(\boldsymbol{x}_i \to \boldsymbol{x}_{i+1}) \cdots p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})\Bigg\}\, d\Xi_i = 0\,. \tag{6.2.14}$$

Ebenso lassen sich die Bedingungen für die Erwartungstreue im Fall $g \equiv 0$ herleiten. Etwas komplizierter ist der Fall, wenn $g(\boldsymbol{x})$ auf einer Teilmenge von $\mathfrak{D}$ verschwindet. Aber auch hier haben die Schwierigkeiten keinen prinzipiellen Charakter, sondern sind rein technischer Art. Wir bemerken noch, daß man die Gleichung (6.2.13) durch verhältnismäßig einfache Umformungen auf die Gestalt

$$\sum_{i=0}^{\infty} \int_{\mathfrak{R}_i} \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)\, q_i h(\boldsymbol{x}_i)$$

$$\times \Bigg\{1 - \sum_{k=i}^{\infty} \int_{\mathfrak{D}} d\boldsymbol{x}_{i+1} \cdots \int_{\mathfrak{D}} d\boldsymbol{x}_k\, p(\boldsymbol{x}_i \to \boldsymbol{x}_{i+1}) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\, g(\boldsymbol{x}_k)\, \omega_i^{(k)}\Bigg\}\, d\Xi_i = 0 \tag{6.2.15}$$

bringen kann, die für die Anwendungen manchmal bequemer ist.

2. Beispiele erwartungstreuer Schätzungen. Wie man leicht an den Gleichungen (6.2.13) und (6.2.15) erkennt, enthält die Klasse der linearen erwartungstreuen Schätzungen auch sehr „pathologische“ Schätzungen, die zu unnötig komplizierten Rechnungen führen. Zur Festlegung der Markoffschen Kette wäre es natürlich, die Dispersionen der Schätzungen unter verschiedenen Voraussetzungen bezüglich der Klasse der zu schätzenden Funktionale und der Klasse von Integralgleichungen miteinander zu vergleichen. So kann leicht der Begriff der zulässigen Schätzung formuliert werden, der sich in diesem Fall in nichts von dem in der mathematischen Statistik üblichen Begriff der Zulässigkeit einer Schätzung unterscheidet. Allerdings ist die analytische Darstellung der Dispersion von Schätzungen für eine genügend umfangreiche Klasse recht schwierig, wie wir noch sehen werden. Man kann das auf die noch unzureichenden Untersuchungen dieses Problems zurückführen. Zweckmäßig wäre es, eine genügend umfangreiche Klasse in numerischer Hinsicht einfacher Schätzungen ausführlicher zu studieren. Die Hervorhebung einer solchen Klasse setzt aber den Beweis voraus, daß es keine weiteren in numerischer Hinsicht einfache Schätzungen gibt, sowie die Präzisierung des Begriffes „einfach“. Man kann erwarten, daß diese Probleme in nächster Zeit die Aufmerksamkeit der Mathematiker finden werden.

Nun geben wir einige einfache Beispiele erwartungstreuer Schätzungen an und untersuchen ihre Dispersionen. Eine der meistgebrauchten Schätzungen ist die folgende:

$$\hat{J}_1[k] = \sum_{i=0}^{k} q_i h(\boldsymbol{x}_i)\,, \qquad g(\boldsymbol{x}) > 0\,, \tag{6.2.16}$$

d. h. $\omega_i^{(k)} = 1$ für alle i und k. Dann ist (6.2.13) offensichtlich erfüllt.

Ebenso einfach überprüft man die Erwartungstreue der Schätzung

$$\hat{J}_2[k] = q_k \frac{h(\boldsymbol{x}_k)}{g(\boldsymbol{x}_k)}, \qquad g(\boldsymbol{x}) > 0\,, \tag{6.2.17}$$

für die

$$\omega_i^{(k)} = \begin{cases} 1/g(\boldsymbol{x}_k) & \text{für} \quad i = k\,, \\ 0 & \text{für} \quad i \neq k \end{cases}$$

gilt. Im Fall $g(\boldsymbol{x}) \equiv 0$ prüft man leicht unmittelbar die Erwartungstreue der Schätzung

$$\hat{J}_1[\infty] = \sum_{i=0}^{\infty} q_i h(\boldsymbol{x}_i) \tag{6.2.18}$$

nach. Setzt man den Ausdruck in den geschweiften Klammern jedes Summanden von (6.2.15) gleich Null, so erhält man hinreichende Bedingungen für die Erwartungstreue in Form eines Gleichungssystems für die Funktionen $\omega_i^{(k)}$. Die Klasse von Schätzungen, für die dieses Gleichungssystem erfüllt ist, wurde von Chissamutdinow [2] betrachtet und von ihm Einsklasse genannt. Wir werden hier einige erwartungstreue Schätzungen anführen, wie sie von Chissamutdinow, aber auch von Michailow [6] hergeleitet wurden. Durch verhältnismäßig einfache Rechnungen kann man überprüfen, daß diese Schätzungen zur Einsklasse gehören

$$\hat{J}_3[k] = \begin{cases} \dfrac{q_{k-l} h(\boldsymbol{x}_{k-l})}{[1 - g(\boldsymbol{x}_{k-l})] \cdots [1 - g(\boldsymbol{x}_{k-1})]\, g(\boldsymbol{x}_k)} & \text{für} \quad k \geqq l\,, \\ 0 & \text{für} \quad k < l\,; \end{cases} \tag{6.2.19}$$

$$\hat{J}_4[k] = \begin{cases} \dfrac{q_{k-l} h(\boldsymbol{x}_{k-l})}{g(\boldsymbol{x}_k)} + \sum\limits_{i=k-l+1}^{k} q_i h(\boldsymbol{x}_i) & \text{für} \quad k \geqq l\,, \\ 0 & \text{für} \quad k < l\,; \end{cases} \tag{6.2.20}$$

$$\hat{J}_5[k] = \begin{cases} \dfrac{q_{k-l} h(\boldsymbol{x}_{k-m})}{[1 - q(\boldsymbol{x}_{k-l-m})] \cdots [1 - q(\boldsymbol{x}_{k-1})]} + \sum\limits_{i=k-m+1}^{k} q_i h(\boldsymbol{x}_i)\,, & l - m \geq 0\,. \\ 0 \end{cases} \tag{6.2.21}$$

Dabei sind l und m beliebige natürliche Zahlen.

So stellen (6.2.19) und (6.2.20) vom Parameter l abhängige Scharen von Schätzungen dar. Die Schätzung (6.2.21) hängt von den beiden Parametern m und l ab.

Eine beliebige Linearkombination erwartungstreuer Schätzungen mit Koeffizienten, deren Summe gleich Eins ist, bildet offensichtlich wieder eine erwartungstreue Schätzung. Insbesondere besteht eine der vielen Verfahren zur Dispersionsverkleinerung in einer geeigneten Wahl dieser Koeffizienten auf Grund einer empirischen Schätzung der Dispersion und Kovarianz im Verlauf der Rechnung.

3. Die Dispersion der Schätzung. Die Frage der Dispersion der Schätzung wird in einer Reihe von Arbeiten behandelt, von denen wir besonders auf die Arbeiten von Solotuchin und Ermakow [1], Solotuchin und Dejew [1] und Michailow [4] hinweisen. Von Maiorow und Frank-Kamenezki [1] wurde kürzlich die Dispersion für eine Klasse von Schätzungen in geschlossener Form erhalten.

Es bereitet offensichtlich keine Schwierigkeit, einen Ausdruck für die Dispersion einer beliebigen Schätzung in der Gestalt eines Integrals über das durch die Markoffsche Kette definierte Maß hinzuschreiben. In manchen Fällen ist es allerdings möglich, einen einfacheren Ausdruck für die Dispersion herzuleiten. Wir beschränken uns auf die Betrachtung der Schätzungen $\hat{J}_1[k]$ und $\hat{J}_2[k]$, die auch am meisten angewendet werden und verhältnismäßig einfach sind. Wir setzen die Konvergenz der im weiteren auftretenden Reihen voraus, und schreiben nun den Ausdruck für die Dispersion der Schätzung $\hat{J}_2[k]$ in folgender Form:

$$\mathsf{D}\hat{J}_2[k] = \sum_{k=0}^{\infty} \int\limits_{\mathfrak{R}_k} \hat{J}_2^2[k]\, \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\, g(\boldsymbol{x}_k)\, d\Xi_k - (\varphi, h)^2$$

$$= \sum_{k=0}^{\infty} \int\limits_{\mathfrak{R}_k} \frac{f^2(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \frac{K^2(\boldsymbol{x}_0, \boldsymbol{x}_1)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1)} \cdots \frac{K^2(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)}{p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)} \frac{h^2(\boldsymbol{x}_k)}{g(\boldsymbol{x}_k)}\, d\Xi_k - (\varphi, h)^2 .$$

Hieraus ergibt sich

$$\mathsf{D}\hat{J}_2[k] = \left(\frac{h^2}{g}, \psi\right) - (\varphi, h)^2 , \tag{6.2.22}$$

wobei ψ eine Lösung der Gleichung

$$\psi(\boldsymbol{x}) = \int\limits_{\mathfrak{D}} \frac{K^2(\boldsymbol{y}, \boldsymbol{x})}{p(\boldsymbol{y} \to \boldsymbol{x})}\, \psi(\boldsymbol{y})\, d\boldsymbol{y} + \frac{f^2(\boldsymbol{x})}{\pi(\boldsymbol{x})} \tag{6.2.23}$$

ist. Führt man die zu (6.2.23) adjungierte Gleichung ein, so kann man das obige Resultat in der Gestalt

$$\mathsf{D}\hat{J}_2[k] = \left(\psi^*, \frac{f^2(\boldsymbol{x})}{\pi(\boldsymbol{x})}\right) - (\varphi, h)^2 \tag{6.2.24}$$

schreiben, wobei ψ^* Lösung der Gleichung

$$\psi^*(\boldsymbol{x}) = \int\limits_{\mathfrak{D}} \frac{K^2(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x} \to \boldsymbol{y})}\, \psi^*(\boldsymbol{y})\, d\boldsymbol{y} + \frac{h^2(\boldsymbol{x})}{g(\boldsymbol{x})} \tag{6.2.25}$$

ist. Gehört also h^2/g zu einem Banachschen Funktionenraum $\mathbf{H}^*$, ist $f^2/\pi \in \mathbf{H}$ und bildet der Operator U mit dem Kern $K^2(\boldsymbol{y}, \boldsymbol{x})/p(\boldsymbol{y} \to \boldsymbol{x})$ den Raum $\mathbf{H}$ in sich ab, so ist für die Existenz und Endlichkeit der Dispersion einer Schätzung $\hat{J}_2[k]$ in dem Raum $\mathbf{H}$ hinreichend und notwendig

$$\lim_{k \to \infty} ||U^k||^{1/k} < 1$$

(vgl. KANTOROWITSCH und AKILOW [1]).

Wir werden die Bedingung nicht weiter konkretisieren, die die Endlichkeit der Dispersion der betrachteten Schätzung sichert. Man prüft in jedem konkreten Fall leicht nach, daß die Menge von Integralgleichungen und entsprechenden Markoffschen Ketten, für die die Dispersion der Schätzung endlich ist, nichtleer ist

Man kann umgekehrt Gleichungen angeben, für die bei beliebiger Wahl der Markoffschen Kette die Norm des Operators U größer als Eins ist. Ein einfaches Beispiel dafür ist das folgende (HALTON [3]). Es sei K ein Matrizenoperator

$$K = \begin{pmatrix} a & -a & 0 \\ a & -a & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \qquad |\lambda| < 1\,, \qquad a > 1\,.$$

Man prüft leicht nach, daß der betragsgrößte Eigenwert von K gleich λ ist. Der entsprechende Operator U wird durch die Matrix

$$\begin{bmatrix} \dfrac{a}{p_{1,1}} & -\dfrac{a}{p_{1,2}} & 0 \\ \dfrac{a}{p_{2,1}} & -\dfrac{a}{p_{2,2}} & 0 \\ 0 & 0 & \dfrac{\lambda}{p_{3,3}} \end{bmatrix},$$

$$0 < p_{i,j} < 1;\quad i, j = 1, 2;\quad 0 < p_{3,3} < 1;\quad p_{1,1} + p_{1,2} \leqq 1;\quad p_{2,1} + p_{2,2} \leqq 1$$

definiert. Bei beliebigem zulässigem $p_{i,j}$ gibt es einen Eigenwert von U, der größer als Eins ist.

Wir kommen nun zur formalen Berechnung der Dispersion der Schätzung $\hat{J}_1[k]$, was gegenüber den bisherigen Betrachtungen wesentlich komplizierter ist. Zunächst berechnen wir $\mathsf{M}\hat{J}_1^2[k]$. Dazu setzt man

$$\pi_k = \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)\,, \qquad \pi_0 = \pi(\boldsymbol{x}_0)\,.$$

Dann gilt

$$\begin{aligned}
\mathsf{M}\hat{J}_1^2[k] &= \sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} d\Xi_k (\hat{J}_1[k])^2\, \pi_k g(\boldsymbol{x}_k) \\
&= \sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} d\Xi_k (\hat{J}_1[k])^2\, \pi_k \left[1 - \int_{\mathfrak{D}} p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})\right] \\
&= \sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} d\Xi_k (\hat{J}_1[k])^2\, \pi_k - \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} (\hat{J}_1[k+1])^2\, \pi_{k+1} \\
&\quad + \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} [(\hat{J}_1[k+1])^2 - (\hat{J}_1[k])^2]\, \pi_{k+1} \\
&= \int_{\mathfrak{D}} d\boldsymbol{x}_0 (\hat{J}_1[0])^2\, \pi_0 + \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} [\hat{J}_1[k+1] + \hat{J}_1[k]]\, [\hat{J}_1[k+1] - \hat{J}_1[k]]\, \pi_{k+1} \\
&= \sum_{k=0}^{\infty} \int_{\mathfrak{R}_k} d\Xi_k q_k^2 h^2(\boldsymbol{x}_k)\, \pi_k + 2 \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} q_{k+1} h(\boldsymbol{x}_{k+1}) \sum_{i=0}^{k} q_i h(\boldsymbol{x}_i)\, \pi_{k+1} \\
&= S_1 + 2S_2\,.
\end{aligned}$$

Man sieht leicht, daß der erste Summand S_1 in dem letzten Ausdruck nichts anderes ist als das Funktional (h^2, ψ), wobei ψ die Lösung der Gleichung (6.2.23) ist. Wir berechnen nun den zweiten Summanden

$$S_2 = \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} q_{k+1} h(\boldsymbol{x}_{k+1}) \sum_{i=0}^{k} q_i h(\boldsymbol{x}_i)\, \pi_{k+1}$$

$$= \sum_{k=0}^{\infty} \int_{\mathfrak{R}_{k+1}} d\Xi_{k+1} \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \cdot \frac{K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_k, \boldsymbol{x}_{k+1})}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})} h(\boldsymbol{x}_{k+1})$$

$$\times \sum_{i=0}^{k} \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \frac{K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)} h(\boldsymbol{x}_i)\, \pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_k \to \boldsymbol{x}_{k+1})$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_{\mathfrak{R}_i} d\Xi_i \frac{f^2(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \frac{K^2(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K^2(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)} h(\boldsymbol{x}_i) \int d\boldsymbol{x}_{i+1} \cdots \int d\boldsymbol{x}_{k+1}$$

$$\times K(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \cdots K(\boldsymbol{x}_k, \boldsymbol{x}_{k+1})\, h(\boldsymbol{x}_{k+1})$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_{\mathfrak{D}} d\boldsymbol{x}\, A_i(\boldsymbol{x})\, B_{k+1}^{(i)}(\boldsymbol{x})$$

mit

$$A_i(\boldsymbol{x}) = \int_{\mathfrak{D}} d\Xi_i \frac{f^2(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \cdot \frac{K^2(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K^2(\boldsymbol{x}_{i-1}, \boldsymbol{x})}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{i-1} \to \boldsymbol{x})} h(\boldsymbol{x}) ,$$

$$B_{k+1}^{(i)}(\boldsymbol{x}) = \int_{\mathfrak{R}_i} d\boldsymbol{x}_{i+1} \cdots d\boldsymbol{x}_{k+1}\, K(\boldsymbol{x}, \boldsymbol{x}_{i+1}) \cdots K(\boldsymbol{x}_k, \boldsymbol{x}_{k+1})\, h(\boldsymbol{x}_{k+1}) .$$

Durch Änderung der Summationsreihenfolge ergibt sich

$$S_2 = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_{\mathfrak{D}} d\boldsymbol{x}\, A_i(\boldsymbol{x})\, B_{k+1}^{(i)}(\boldsymbol{x}) = \sum_{i=0}^{\infty} \sum_{k=i}^{\infty} \int_{\mathfrak{D}} d\boldsymbol{x}\, A_i(\boldsymbol{x})\, B_{k+1}^{(i)}(\boldsymbol{x}) .$$

Nun hängt $B_{k+1}^{(i)}(\boldsymbol{x})$ nur von der Differenz der Indizes $s = k - i$ ab. Ändert man die Bezeichnung in $B_{k+1}^{(i)}(\boldsymbol{x})$ etwas ab und summiert nach der Veränderlichen s, so erhält man

$$S_2 = \lim_{N \to \infty,\, N_1 \to \infty} \sum_{i=0}^{N} \sum_{k=i}^{N_1} \int_{\mathfrak{D}} d\boldsymbol{x}\, A_i(\boldsymbol{x})\, B_{k+1}^{(i)}(\boldsymbol{x})$$

$$= \lim_{N \to \infty,\, N_1 \to \infty} \sum_{i=0}^{N} \sum_{s=0}^{N_1 - i} \int_{\mathfrak{D}} d\boldsymbol{x}\, A_i(\boldsymbol{x})\, B_{s+1}(\boldsymbol{x})$$

mit

$$B_s(\boldsymbol{x}) = \int_{\mathfrak{D}} d\boldsymbol{y}_1 \cdots \int_{\mathfrak{D}} d\boldsymbol{y}_s\, K(\boldsymbol{x}, \boldsymbol{y}_1) \cdots K(\boldsymbol{y}_{s-1}, \boldsymbol{y}_s)\, h(\boldsymbol{x}_s) .$$

Nun gilt aber

$$\sum_{s=0}^{\infty} B_{s+1}(\boldsymbol{x}) = \int_{\mathfrak{D}} K(\boldsymbol{x}, \boldsymbol{y})\, \varphi^*(\boldsymbol{y})\, d\boldsymbol{y} \qquad \text{und} \qquad \sum_{i=0}^{\infty} A_i(\boldsymbol{x}) = h(\boldsymbol{x})\, \psi(\boldsymbol{x}) ,$$

wobei $\varphi^*(\boldsymbol{x})$ und $\psi(\boldsymbol{x})$ Lösungen von (6.2.4) bzw. (6.2.23) sind.

Damit erhält man

$$S_2 = \int_{\mathcal{D}} d\boldsymbol{x}\, h(\boldsymbol{x})\, \psi(\boldsymbol{x}) \int_{\mathcal{D}} K(\boldsymbol{x}, \boldsymbol{y})\, \varphi^*(\boldsymbol{y})\, d\boldsymbol{y}$$

$$= \int_{\mathcal{D}} d\boldsymbol{x}\, h(\boldsymbol{x})\, \psi(\boldsymbol{x})\, [\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})]\, d\boldsymbol{x} \tag{6.2.26}$$

und schließlich

$$\mathsf{D}\hat{J}_1[k] = (2h\varphi^* - h^2, \psi) - (\varphi, h)^2 \,. \tag{6.2.27}$$

Genau wie bei der Schätzung $\hat{J}_2[k]$ kann man zur adjungierten Gleichung übergehen und erhält

$$\mathsf{D}\hat{J}_1 = \left(\psi^*, \frac{f^2}{\pi}\right) - (\varphi, h)^2 \,, \tag{6.2.28}$$

wobei ψ^* hier Lösung der Gleichung

$$\psi^*(\boldsymbol{x}) = \int_{\mathcal{D}} \frac{K^2(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x} \to \boldsymbol{y})}\, \psi^*(\boldsymbol{y})\, d\boldsymbol{y} + h(\boldsymbol{x})\, \big(2\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})\big) \tag{6.2.29}$$

ist. Durch völlig analoge Überlegungen kann man die Dispersion der Schätzung $\hat{J}_1[\infty]$ berechnen. Dabei hat der Ausdruck für die Dispersion dieselbe Gestalt $\mathsf{D}\hat{J}_1[\infty] = (2h\varphi^* - h^2, \psi) - (h, \varphi)^2$, und es gilt auch die adjungierte Formel (6.2.28).

Die erhaltenen Resultate bilden den Inhalt des folgenden Satzes.

Satz 6.3. *Wenn die Existenz und Endlichkeit der Dispersion vorausgesetzt wird, dann gilt:*

1. *Die Dispersion der Schätzung* $\hat{J}_1[k]$ *bzw.* $\hat{J}_1[\infty]$ *des Funktionals* J *ergibt sich aus* (6.2.27) *und* (6.2.28), *wobei* ψ *und* ψ^* *Lösungen der Gleichungen* (6.2.23) *bzw.* (6.2.29) *sind.*

2. *Die Dispersion der Schätzung* $\hat{J}_2[k]$ *ergibt sich aus* (6.2.22) *und* (6.2.24), *wobei* ψ *und* ψ^* *Lösungen der Gleichungen* (6.2.23) *bzw.* (6.2.25) *sind.*

Bemerkung. Es sei $r(\boldsymbol{x})$ eine beliebige streng positive und in $\mathcal{D}$ endliche Funktion. Wegen

$$\int_{\mathfrak{R}_k} \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \frac{K^2(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K^2(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)}\, h(\boldsymbol{x}_k)\, d\Xi_k$$

$$= \int_{\mathfrak{R}_k} \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)}\, r(\boldsymbol{x}_0)\, \frac{K^2(\boldsymbol{x}_0, \boldsymbol{x}_1)}{p(\boldsymbol{x}_0 \to \boldsymbol{x}_1)} \frac{r(\boldsymbol{x}_1)}{r(\boldsymbol{x}_0)} \cdots \frac{K^2(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)}{p(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)} \frac{r(\boldsymbol{x}_k)}{r(\boldsymbol{x}_{k-1})} \frac{h(\boldsymbol{x}_k)}{r(\boldsymbol{x}_k)}\, d\Xi_k \tag{6.2.30}$$

kann man die Bedingung $\lim\limits_{k\to\infty} ||U^k||^{1/k} < 1$ durch die dazu äquivalente Bedingung $\lim\limits_{k\to\infty} ||\tilde{U}^k||^{1/k} < 1$ ersetzen, wobei $\tilde{U}$ der zum Kern $K^2(\boldsymbol{y}, \boldsymbol{x})\, r(\boldsymbol{x})/p(\boldsymbol{y} \to \boldsymbol{x})\, r(\boldsymbol{y})$ gehörige Operator ist.

Ein Vergleich der Ausdrücke für die Dispersionen der Schätzungen $\hat{J}_1[k]$ und $\hat{J}_2[k]$ zeigt, daß in Abhängigkeit von der Gestalt der Gleichung und von der benutzten Markoffschen Kette entweder $\hat{J}_1[k]$ oder $\hat{J}_2[k]$ günstiger sein kann, d. h., man kann nicht behaupten, daß eine der beiden Schätzungen von vornherein schlechter ist als die andere. Zur Begründung dieser Behauptung genügt es, die Funktionale $(h^2/g, \psi)$ und $(2h\varphi^* - h^2, \psi)$ zu vergleichen. ψ ist in beiden Fällen dieselbe Funktion. Wir nehmen nun an, daß h, φ^* und $K(\boldsymbol{x}, \boldsymbol{y})$ positive Funktionen sind. Dann gilt

$$\varphi^* = K^*\varphi^* + h \geqq h \qquad \text{für alle } \boldsymbol{x}$$

und daher $2h\varphi^* - h^2 \geqq h^2$. Weiter gibt es ein solches K, daß

$$2h\varphi^* - h^2 = h^2 + \delta$$

mit $\delta = \delta(\boldsymbol{x}) > 0$ erfüllt ist. In diesem Fall ist

$$(2h\varphi^* - h^2, \psi) = (h^2 + \delta, \psi) ,$$

und für

$$g(\boldsymbol{x}) > \frac{h^2(\boldsymbol{x})}{h^2(\boldsymbol{x}) + \delta(\boldsymbol{x})}$$

gilt die Ungleichung

$$\left(\frac{h^2}{g}, \psi\right) < (2h\varphi^* - h^2, \psi) .$$

Ist aber $g(\boldsymbol{x})$ genügend klein, so gilt in der letzten Ungleichung offensichtlich das entgegengesetzte Zeichen.

Der Rechenaufwand bei Anwendung der Schätzungen $\hat{J}_1$ und $\hat{J}_2$ unterscheidet sich bei einfachen $h(\boldsymbol{x})$ nur wenig. Die zusätzliche Arbeit der Summation bei Anwendung der Schätzung $\hat{J}_1$ kann gewöhnlich vernachlässigt werden. Wenn aber die Berechnung der Werte von h sehr aufwendig ist, dann kann die Schätzung $\hat{J}_2$ bei entsprechender Wahl von $g(\boldsymbol{x})$ in einigen Fällen günstiger sein. Dabei sollte man beachten, daß sich diese Aussagen auf die Berechnung eines konkreten Funktionals beziehen.

Wir untersuchen nun die Form der Dispersion in einigen konkreten Fällen, die von besonderem Interesse sein könnten:

a) $K(\boldsymbol{y}, \boldsymbol{x})$ ist eine nichtnegative Funktion, die der Bedingung

$$\int_{\mathfrak{D}} K(\boldsymbol{y}, \boldsymbol{x})\, d\boldsymbol{x} = 1 - g(\boldsymbol{y}) , \qquad 0 \leqq g(\boldsymbol{y}) < 1 ,$$

genügt, für $f(\boldsymbol{x})$ gilt $f(\boldsymbol{x}) \geqq 0$. Hier kann man $K(\boldsymbol{y}, \boldsymbol{x}) \equiv p(\boldsymbol{y} \to \boldsymbol{x})$ und $\pi(\boldsymbol{x}) = f(\boldsymbol{x}) / \int_{\mathfrak{D}} f(\boldsymbol{x})\, d\boldsymbol{x}$ setzen. Wir nehmen weiterhin an, daß $\int f(\boldsymbol{x})\, d\boldsymbol{x} = 1$ ist, da das stets durch eine einfache Variablensubstitution erreichbar ist. Dann gilt

$$\left.\begin{aligned} &\psi(\boldsymbol{x}) \equiv \varphi(\boldsymbol{x}) , \qquad \mathsf{D}\hat{J}_2 = \left(\frac{h^2}{g}, \varphi\right) - (\varphi, h)^2 , \\ &\mathsf{D}\hat{J}_1 = (2h\varphi^* - h^2, \varphi) - (\varphi, h)^2 . \end{aligned}\right\} \qquad (6.2.31)$$

b) $p(\boldsymbol{y} \to \boldsymbol{x})$ ist im Gebiet $\mathfrak{D}$ konstant und auch $\pi(\boldsymbol{x})$ ist eine Konstante. Das ist ein ausgearteter Fall, denn $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ stellen unabhängige und im Gebiet $\mathfrak{D}$ gleichverteilte Zufallsgrößen dar. Das Rechenverfahren entspricht der Berechnung jeden Gliedes der numerischen Reihe für φ nach der einfachsten Formel (4.1.18) (gewöhnliche Monte-Carlo-Methode). Die einzige Besonderheit bei der Berechnung des folgenden Gliedes der Reihe ist die Benutzung der früher gewählten $\boldsymbol{x}_i$-Werte, die bei der Berechnung des vorangehenden Gliedes der Reihe benutzt wurden. Die Funktion ψ genügt der Gleichung

$$\psi(\boldsymbol{x}) = \int_{\mathfrak{D}} K^2(\boldsymbol{y}, \boldsymbol{x})\, \psi(\boldsymbol{y})\, d\boldsymbol{y} + f^2(\boldsymbol{x})\,.$$

c) Der Kern $K(\boldsymbol{y}, \boldsymbol{x})$ und die Funktion f ändern in $\mathfrak{D}$ ihr Vorzeichen, aber für p und π kann man $|K(\boldsymbol{y}, \boldsymbol{x})|$ und $|f(\boldsymbol{x})|$ wählen. Die Funktion ψ genügt der Gleichung

$$\psi(\boldsymbol{x}) = \int_{\mathfrak{D}} |K(\boldsymbol{y}, \boldsymbol{x})|\, \psi(\boldsymbol{y})\, d\boldsymbol{y} + |f(\boldsymbol{x})|\,.$$

Die Dispersionen der Schätzungen (6.2.19) bis (6.2.21) werden wir nicht in geschlossener Form angeben. Für die Schätzung $\hat{J}_3$ ist das übrigens ebenso wie für $\hat{J}_1$ nicht besonders schwierig.

4. Über das Verschwinden der Dispersion einer Schätzung. Es sei $\hat{J}$ eine erwartungstreue Schätzung des Funktionals J und $\{\pi, p\}$ eine Markoffsche Kette, die den Bedingungen (6.2.6) genügt. Die Wahl einer anderen Kette $\{\tilde{\pi}, \tilde{p}\}$, für die auch die Bedingungen (6.2.6) erfüllt sein sollen, bedeutet den Übergang auf ein anderes Maß. Mit anderen Worten, ist Ξ_1 eine feste Trajektorie der Kette $\{\pi, p\}$ und μ_1 das Maß auf der Menge solcher Trajektorien, und sind Ξ_2 und μ_2 Trajektorie bzw. Maß in bezug auf $\{\tilde{\pi}, \tilde{p}\}$, dann gilt

$$J = \int \hat{J} \mu_1(d\Xi_1) = \int \hat{J}\, \frac{d\mu_1}{d\mu_2}\, \mu_2(d\Xi_2)\,. \tag{6.2.32}$$

Die absolute Stetigkeit des Maßes μ_1 bezüglich μ_2 ist gesichert, wenn die Bedingungen (6.2.6) erfüllt sind. Aus (6.2.32) folgt vor allem, daß lineare Schätzungen bei Änderung des Maßes ihre Form behalten. Außerdem kann man nach den allgemeinen Resultaten (Kap. IV, § 1) ein Maß $\mu = \mu_0$ derart wählen, daß die Dispersion der Schätzung $\hat{J} d\mu_1/d\mu_0$ verschwindet.

Die formale Anwendung der Beziehung (4.2.3) führt zu folgender Aussage über das Maß μ_0 im Fall der Schätzung $\hat{J}_2[k]$: Einer Trajektorie der Länge k entspricht die Wahrscheinlichkeitsdichte

$$\pi_k^0 = f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots K(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)\, \frac{h(\boldsymbol{x}_k)}{J} \tag{6.2.33}$$

(wobei die Funktionen f, K und h in $\mathfrak{D}$ als nichtnegativ vorausgesetzt werden). Wie auch bei der Integralberechnung wird für die Anwendung ein Maß gesucht, das zu einer genügend kleinen Dispersion führt. Vom numerischen Standpunkt ist es bequem, eine homogene Markoffsche Kette zu konstruieren, so daß der Mittelwert über deren Trajektorien eine kleine Dispersion besitzt. Die Konstruktion

einer solchen Kette, die auch der gegebenen Aufgabe entspricht, ist nicht trivial. Unter Benutzung der Lösung φ^* von Gleichung (6.2.4) kann man in diesem Fall folgendermaßen vorgehen. Man schreibt (6.2.33) in der Form

$$\pi_k^0 = \frac{f(\boldsymbol{x}_0)\,\varphi^*(\boldsymbol{x}_0)}{J}\, K(\boldsymbol{x}_0, \boldsymbol{x}_1)\,\frac{\varphi^*(\boldsymbol{x}_1)}{\varphi^*(\boldsymbol{x}_0)} \times K(\boldsymbol{x}_1, \boldsymbol{x}_2)\,\frac{\varphi^*(\boldsymbol{x}_2)}{\varphi^*(\boldsymbol{x}_1)} \cdots K(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k)\,\frac{\varphi^*(\boldsymbol{x}_k)}{\varphi^*(\boldsymbol{x}_{k-1})} \cdot \frac{h(\boldsymbol{x}_k)}{\varphi^*(\boldsymbol{x}_k)}.$$

Aus (6.2.4) und der Positivität der Funktionen K und h folgt nämlich $\varphi^*(\boldsymbol{x}) \geqq h(\boldsymbol{x})$, und man kann $h(\boldsymbol{x})/\varphi^*(\boldsymbol{x}) = g(\boldsymbol{x})$ setzen. Außerdem gilt

$$\varphi^*(\boldsymbol{y}) - h(\boldsymbol{y}) = \int_{\mathcal{D}} K(\boldsymbol{y}, \boldsymbol{x})\,\varphi^*(\boldsymbol{x})\,d\boldsymbol{x}$$

und

$$K(\boldsymbol{y}, \boldsymbol{x})\,\frac{\varphi^*(\boldsymbol{x})}{\varphi^*(\boldsymbol{y})} = \frac{K(\boldsymbol{y}, \boldsymbol{x})\,\varphi^*(\boldsymbol{x})}{\varphi^*(\boldsymbol{y}) - h(\boldsymbol{y})}\left[1 - \frac{h(\boldsymbol{y})}{\varphi^*(\boldsymbol{y})}\right] = \frac{K(\boldsymbol{y}, \boldsymbol{x})\,\varphi^*(\boldsymbol{x})}{\int K(\boldsymbol{y}, \boldsymbol{x})\,\varphi^*(\boldsymbol{x})\,d\boldsymbol{x}}\,[1 - g(\boldsymbol{y})]\,,$$

d. h., man kann

$$K(\boldsymbol{y}, \boldsymbol{x})\,\frac{\varphi^*(\boldsymbol{x})}{\varphi^*(\boldsymbol{y})} = p(\boldsymbol{y} \to \boldsymbol{x})$$

setzen. Weiterhin ist es klar, daß man $f(\boldsymbol{x})\,\varphi^*(\boldsymbol{x})/J$ als $\pi(\boldsymbol{x})$ wählen kann, da die Normierungsbedingung $(f, \varphi^*) = J$ erfüllt ist. Das Resultat kann man im folgenden Satz formulieren.

Satz 6.4. *Die Funktionen $f(\boldsymbol{x})$, $h(\boldsymbol{x})$ und $K(\boldsymbol{x}, \boldsymbol{y})$ seien nichtnegativ für alle $\boldsymbol{x}$ und $\boldsymbol{y}$ aus $\mathcal{D}$. Wählt man dann die Markoffsche Kette*

$$\left\{f(\boldsymbol{x})\,\frac{\varphi^*(\boldsymbol{x})}{J},\ K(\boldsymbol{y}, \boldsymbol{x})\,\frac{\varphi^*(\boldsymbol{x})}{\varphi^*(\boldsymbol{y})}\right\},$$

so ist die Dispersion der Schätzung $\hat{J}_2[k]$ gleich Null.

Zur Realisierung der Markoffschen Kette mit verschwindender Dispersion benötigt man also außer der Größe des Funktionals J auch die Funktion $\varphi^*(\boldsymbol{x})$. In vielen physikalischen Aufgaben hat die Funktion $\varphi^*(\boldsymbol{x})$ einen bestimmten Sinn, und man kann begründete Aussagen über ihr Verhalten machen, d. h., $\varphi^*(\boldsymbol{x})$ kann näherungsweise gegeben sein. Das Ergebnis von Satz 6.4 ist dabei offensichtlich nur dann von Interesse, wenn die Dispersion der Schätzung in bezug auf kleine Änderungen von $\varphi^*(\boldsymbol{x})$ stabil ist. Man kann zeigen, daß diese Stabilität tatsächlich gilt. Wir ersetzen dazu φ^* durch $\hat{\varphi}^*(\boldsymbol{x})$ mit

$$\hat{\varphi}^*(\boldsymbol{x}) = \varphi^*(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})\,, \qquad \varepsilon(\boldsymbol{x}) \in \mathrm{F}^*\,, \qquad \hat{\varphi}^*(\boldsymbol{x}) > 0\,,$$

und setzen

$$\pi(\boldsymbol{x}) = \frac{f(\boldsymbol{x})\,\hat{\varphi}^*(\boldsymbol{x})}{J + (f, \varepsilon)}\,, \qquad p(\boldsymbol{y} \to \boldsymbol{x}) = \frac{K(\boldsymbol{y}, \boldsymbol{x})\,\hat{\varphi}^*(\boldsymbol{x})}{\hat{\varphi}^*(\boldsymbol{y})}\,.$$

Dann erhält man nach einfacher Zwischenrechnung

$$g(\boldsymbol{y}) = \frac{1}{\hat{\varphi}^*(\boldsymbol{y})} \left[h(\boldsymbol{y}) + \varepsilon(\boldsymbol{y}) - \int K(\boldsymbol{y}, \boldsymbol{x})\, \varepsilon(\boldsymbol{x})\, d\boldsymbol{x}\right].$$

Ist dabei $0 < g(\boldsymbol{y}) < 1$, so ergibt sich $\psi(\boldsymbol{x})$ aus der Gleichung (6.2.4)

$$\psi(\boldsymbol{x}) = \frac{\varphi(\boldsymbol{x})}{\hat{\varphi}^*(\boldsymbol{x})} [J + (f, \varepsilon)]$$

und nach Satz 6.3

$$\mathbf{D}\hat{J}_2 = \left(\frac{h^2}{g}, \psi\right) - J^2,$$

wobei die Möglichkeit der Anwendung von Satz 6.3 durch die Bemerkung zu diesem Satz begründet werden kann. Wie man leicht sieht, strebt mit $||\varepsilon||_{\mathbf{F}^*} \to 0$ auch $\mathbf{D}\hat{J}_2$ gegen Null.

Unsere bisherigen Betrachtungen bezogen sich auf den Fall einer positiven Absorptionswahrscheinlichkeit und der einfachsten Schätzung $\hat{J}_2[k]$. Dieser Fall ist in genügendem Maß charakteristisch. Für $g(\boldsymbol{x}) \equiv 0$ kann man auch eine Markoffsche Kette konstruieren, für die die Dispersion der Schätzung $\hat{J}_1[\infty]$ verschwindet. Es gilt hier folgendes Resultat.

Satz 6.5. *Die Funktionen $f(\boldsymbol{x})$, $h(\boldsymbol{x})$ und $K(\boldsymbol{x}, \boldsymbol{y})$ seien positiv. Setzt man dann*

$$\pi(\boldsymbol{x}) = \frac{f(\boldsymbol{x})\, \varphi^*(\boldsymbol{x})}{J}$$

und

$$p(\boldsymbol{y} \to \boldsymbol{x}) = \frac{K(\boldsymbol{y}, \boldsymbol{x})\, \varphi^*(\boldsymbol{x})}{\int K(\boldsymbol{y}, \boldsymbol{x})\, \varphi^*(\boldsymbol{x})\, d\boldsymbol{x}},$$

so ist die Dispersion von $\hat{J}_1[\infty]$ gleich Null.

Beweis. In diesem Fall ist $g(\boldsymbol{y}) \equiv 0$. Die Normierung von $\pi(\boldsymbol{x})$ und $p(\boldsymbol{y} \to \boldsymbol{x})$ überprüft man leicht direkt. Wir benutzen nun den Ausdruck (6.2.28) für $\mathbf{D}\hat{J}_1[\infty]$ und finden dazu $\psi^*(\boldsymbol{x})$ aus der Gleichung

$$\psi^*(\boldsymbol{x}) = \int \frac{K^2(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x} \to \boldsymbol{y})}\, \psi^*(\boldsymbol{y})\, d\boldsymbol{y} + h(\boldsymbol{x})\, [2\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})].$$

Wegen

$$p(\boldsymbol{x} \to \boldsymbol{y}) = \frac{K(\boldsymbol{x}, \boldsymbol{y})\, \varphi^*(\boldsymbol{y})}{\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})}$$

gilt

$$\psi^*(\boldsymbol{x}) = \int \frac{K(\boldsymbol{x}, \boldsymbol{y})\, [\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})]}{\varphi^*(\boldsymbol{y})}\, \psi^*(\boldsymbol{y})\, d\boldsymbol{y} + h(\boldsymbol{x})\, [2\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})]. \tag{6.2.34}$$

Diese formale Rechnung findet wie schon früher ihre Begründung durch die Bemerkung zu Satz 6.3.

Die Gleichung (6.2.34) schreiben wir nun in der Gestalt

$$\psi^*(\boldsymbol{x}) = [\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})] \int K(\boldsymbol{x}, \boldsymbol{y}) \frac{\psi^*(\boldsymbol{y})}{\varphi^*(\boldsymbol{y})} d\boldsymbol{y} + h(\boldsymbol{x}) [2\varphi^*(\boldsymbol{x}) - h(\boldsymbol{x})]$$

oder in der dazu äquivalenten Form

$$\psi^*(\boldsymbol{x}) = \varphi^*(\boldsymbol{x}) \left[\int K(\boldsymbol{x}, \boldsymbol{y}) \frac{\psi^*(\boldsymbol{y})}{\varphi^*(\boldsymbol{y})} d\boldsymbol{y} + 2h(\boldsymbol{x}) \right] - h(\boldsymbol{x}) \left[\int K(\boldsymbol{x}, \boldsymbol{y}) \frac{\psi^*(\boldsymbol{y})}{\varphi^*(\boldsymbol{y})} d\boldsymbol{y} + h(\boldsymbol{x}) \right].$$

Wie man leicht sieht, ist $\psi^*(\boldsymbol{x}) = [\varphi^*(\boldsymbol{x})]^2$ eine Lösung dieser Gleichung. Setzt man in den Ausdruck (6.2.28) für $\mathbf{D}\hat{J}_1[\infty]$ die Werte für $\psi^*(\boldsymbol{x})$ und $\pi(\boldsymbol{x})$ ein, so ergibt sich

$$\mathbf{D}\hat{J}_1[\infty] = \left((\varphi^*)^2, \frac{fJ}{\varphi^*} \right) - J^2 = 0 ,$$

was zu beweisen war.

Michailow [4] erhält in einer Arbeit das stärkere Ergebnis, daß unter den Bedingungen von Satz 6.5 die Schätzung $\hat{J}_1[\infty]$ auf allen Trajektorien der entsprechenden Markoffschen Kette konstant ist. Das wird mit Methoden bewiesen, die den beim Beweis von Satz 6.1 in § 1 gebrauchten Methoden ähnlich sind.

Ebenso wie für die Schätzung $\hat{J}_2[k]$ kann man zeigen, daß die Dispersion der Schätzung $\hat{J}_1$ für $\hat{\varphi}^*(\boldsymbol{x}) = \varphi^*(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})$ anstelle von $\varphi^*(\boldsymbol{x})$ der Norm nach bei kleinem ε stabil ist. Die Rechnungen sind in diesem Fall allerdings komplizierter.

5. Weitere Methoden zur Dispersionsverkleinerung. Wie wir schon sahen, gibt es bei der Integration über die Trajektorien einer Markoffschen Kette gegenüber den allgemeinen Resultaten des vierten Kapitels gewisse Besonderheiten, insbesondere bei der Methode der wesentlichen Stichprobe. Was die Methode der Abtrennung des Hauptteils betrifft, so bereitet ihre Anwendung keine Schwierigkeiten. Es sei nämlich $\varphi = K\varphi + f$ die zu lösende Gleichung und $\varphi' = K'\varphi' + f'$ eine dazu ähnliche Gleichung mit bekannter Lösung. Dann ist die Differenz $\varphi' - \varphi = \Delta\varphi$ Lösung der Gleichung

$$\Delta\varphi = K\Delta\varphi + (K' - K)\,\varphi' + f' - f ,$$

die sich nur in der rechten Seite von der Ausgangsgleichung unterscheidet. Ist $\Delta\varphi$ klein, so kann der Gewinn bedeutend sein. Keiner besonderen Erklärung bedarf der Fall, daß einige der ersten Glieder der Neumannschen Reihe analytisch berechnet werden können oder daß die Integration nach einigen Veränderlichen leicht analytisch durchführbar ist. Damit kann man in der Regel den Rechenaufwand wesentlich verkleinern.

Mit den stochastischen Quadraturformeln kann man den Übergang zu einem verzweigten Prozeß verbinden (Frolow und Tschenzow [2], Frolow [1]).

Wir stellen den Kern $K(\boldsymbol{y}, \boldsymbol{x})$ in Form einer Summe

$$K(\boldsymbol{y}, \boldsymbol{x}) = \sum_{j=0}^{m-1} K_j(\boldsymbol{y}, \boldsymbol{x}) \tag{6.2.35}$$

dar und entsprechend den Operator K durch

$$K = \sum_{j=0}^{m-1} K_j \,.$$

Das l-te Glied der Neumannschen Reihe der Gleichung (6.2.1) kann in der Form

$$\varphi_l = K^l f = [K_1 + K_2 + \cdots + K_m]^l f = \sum_{i_1, \ldots, i_l = 0}^{m-1} K_{i_1} \cdots K_{i_l} f \tag{6.2.36}$$

dargestellt werden. Es ist bequem, hier eine symbolische Bezeichnung einzuführen. Dazu faßt man $i_1, \ldots, i_l$ als Zahl aus einem m-Positionssystem auf. Bezeichnet man nun mit Γ_l die Menge aller l-stelligen Zahlen des m-Positionssystem und mit $\gamma_{(l)} = (i_1, \ldots, i_l)$ eine bestimmte Zahl daraus, so kann man

$$\varphi_l = \sum_{\gamma \in \Gamma_l} K_{i_1} \cdots K_{i_l} f$$

schreiben. Nun sei $p_i(\boldsymbol{y} \to \boldsymbol{x})$, $i = 1, \ldots, n$, die Übergangsdichte einer Markoffschen Kette, wobei $p_i(\boldsymbol{y} \to \boldsymbol{x}) > 0$ für $K_i(\boldsymbol{y}, \boldsymbol{x}) > 0$ und

$$\int_{\mathfrak{D}} p_i(\boldsymbol{y} \to \boldsymbol{x}) \, d\boldsymbol{x} = 1 - g_i(\boldsymbol{y}) \,, \qquad 0 \leqq g_i(\boldsymbol{y}) < 1$$

gilt.

Wir konstruieren nun einen Verzweigungsprozeß, in dem jedes Teilchen m analoge Teilchen bildet, die den Index j, $1 \leqq j \leqq m$, erhalten. Das im Punkt $\boldsymbol{y}$ entstandene Teilchen mit dem Index j geht dann entweder mit einer Wahrscheinlichkeit, deren Dichte $p_j(\boldsymbol{y} \to \boldsymbol{x})/\big(1 - g_j(\boldsymbol{y})\big)$ beträgt, in den Punkt $\boldsymbol{x}$ über, wobei sich seine Masse um den Faktor $K_j(\boldsymbol{y}, \boldsymbol{x})/p_j(\boldsymbol{y} \to \boldsymbol{x})$ vervielfacht, oder es wird mit der Wahrscheinlichkeit $g_j(\boldsymbol{y})$ im Punkt $\boldsymbol{y}$ absorbiert. Die auf diese Weise entstandene Teilchenlawine wird durch ein Teilchen hervorgerufen, dessen Teilung mit seiner Entstehung beginnt. Die Wahrscheinlichkeitsdichte, die einem Zweig des Trajektorienbaumes von der Länge l entspricht, beträgt

$$\pi(\boldsymbol{x}_0) \, p_{i_1}(\boldsymbol{x}_0 \to \boldsymbol{x}_1^{(i_1)}) \, p_{i_2}(\boldsymbol{x}_1^{(i_1)} \to \boldsymbol{x}_2^{(i_2)}) \cdots p_{i_l}(\boldsymbol{x}_{l-1}^{(i_{l-1})} \to \boldsymbol{x}_l^{(i_l)}) \, g_{i_l}(\boldsymbol{x}_l^{(i_l)}) \,.$$

Hierbei ist i_l der Index des Teilchens selbst, und $i_{l-1}, i_{l-2}, \ldots, i_1$ sind die Indizes seiner „Vorfahren" in den vorangehenden Generationen. Dann gilt der folgende Satz

Satz 6.6. *Es sei*

$$\hat{J} = \sum_{i=1}^{k} q_i \omega_i^{(k)}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k) \, h(\boldsymbol{x}_i)$$

eine lineare erwartungstreue (*nach* (6.2.15)) *Schätzung auf den Trajektorien der Kette* $\{\pi, p\}$. *Dann ist die lineare Schätzung*

$$\hat{J} = \sum_{i=1}^{k} \tilde{q}_i \omega_i^{(k)}(\boldsymbol{x}_0, \boldsymbol{x}_1^{(i_1)}, \ldots, \boldsymbol{x}_l^{(i_l)}) \, h(\boldsymbol{x}_l^{(i_l)}) \tag{6.2.37}$$

auf den Trajektorien des Verzweigungsprozesses, für die

$$p(\boldsymbol{y} \to \boldsymbol{x}) = \sum_{i=1}^{m} s_i p_i(\boldsymbol{y} \to \boldsymbol{x}) \,, \qquad s_i > 0 \,, \qquad \sum_{i=1}^{m} s_i = 1 \,,$$

$$\tilde{q}_i = \frac{f(\boldsymbol{x}_0)}{\pi(\boldsymbol{x}_0)} \cdot \frac{K_{i_1}(\boldsymbol{x}_0, \boldsymbol{x}_1^{(i_1)})}{p_{i_1}(\boldsymbol{x}_0 \to \boldsymbol{x}_1^{(i_1)})} \cdots \frac{K_{i_i}(\boldsymbol{x}_{i-1}^{(i_{i-1})}, \boldsymbol{x}_i^{(i_i)})}{p_{i_i}(\boldsymbol{x}_{i-1}^{(i_{i-1})} \to \boldsymbol{x}_i^{(i_i)})}$$

gilt und die Bedingungen (6.2.6) *und* (6.2.36) *erfüllt sind, eine erwartungstreue Schätzung des Funktionals* J.

Der Beweis erfolgt unmittelbar aus (6.2.36), aus der Linearität der Schätzung und aus der vorausgesetzten Gültigkeit von (6.2.15). Daher kann man außer einer homogenen Markoffschen Kette zur Lösung von Integralgleichungen auch eine verzweigte Kette benutzen.

Gibt es ein durch eine homogene Markoffsche Kette beschriebenes Modell einer Erscheinung, dann kann man in gewissen Fällen zur Verkleinerung der Dispersion der gesuchten Größe unter Benutzung von Satz 6.6 zu einem verzweigten Prozeß übergehen. Umgekehrt gibt derselbe Satz die Möglichkeit, zu einer linearen Kette überzugehen, wenn es ein durch eine verzweigte Kette beschriebenes Modell gibt. Die Anwendung einer stochastischen Quadraturformel bezüglich jeder Veränderlichen $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots$ liefert auch eine Zerlegung (6.2.35) des ursprünglichen Kernes in eine Summe von Kernen. Im allgemeinen können allerdings beim Gebrauch einer Quadraturformel die eine Markoffsche Kette auszeichnenden Eigenschaften verloren gehen.

Numerische Probleme, die mit der Modellierung verzweigter Prozesse verknüpft sind, wurden in der Literatur recht ausführlich untersucht (Golenko [1], Frolow [1]). Wenn wir als einen Versuch die Realisierung des ganzen Baumes ansehen (wobei der positive Kern $K(\boldsymbol{y}, \boldsymbol{x})$ in positive Summanden $K_j(\boldsymbol{y}, \boldsymbol{x})$ zerlegt wird), dann führt die Verzweigung zur Dispersionsverkleinerung eines Versuches aber zur Vergrößerung des Rechenaufwandes. Unter diesen Bedingungen ist die Herleitung numerischer Verfahren hoher Effektivität eine wichtige und noch verhältnismäßig wenig untersuchte Aufgabe. Für das Problem des Strahlendurchgangs wurde diese Frage in einer Reihe von Arbeiten untersucht (z. B. Kahn [1], Ogibin [1]).

Abschließend soll noch einmal bemerkt werden, daß die bevorzugte Betrachtung homogener Markoffscher Prozesse ausschließlich im Hinblick auf die einfache Darstellung geschah. Tatsächlich bleiben lineare Schätzungen, für die die $\omega_i^{(k)}$ den Bedingungen (6.2.15) genügen, auch bei nichthomogenen Markoffschen Prozessen und sogar bei Prozessen allgemeinerer Art erwartungstreu, wenn nur die q_i in entsprechender Weise gewählt werden. Dieselbe Bemerkung gilt auch für verzweigte Prozesse, bei denen man einen Zerfall des Teilchens in eine bei jedem Schritt verschiedene Anzahl von Teilchen betrachten kann (entsprechend der Zerlegung des Kerns in Summanden). So kann man jedem Markoffschen Modell eine unendliche Menge von Modellen zuordnen, die im allgemeinen zu einem unterschiedlichen Arbeitsaufwand bei der Lösung der gegebenen Aufgabe führen. Das grundlegende Problem besteht daher in der Wahl eines konkreten und effektiven

Modells. Dieses Problem ist aber in der gegenwärtigen Zeit noch nicht einmal vollständig formuliert. Meistens wird es durch intuitive (oder in Aufgaben mit physikalischem Inhalt durch sogenannte „physikalische") Überlegungen gelöst. Die Auswahl des geeigneten Modells kann auch experimentell auf der Grundlage einer vorbereitenden kleinen Stichprobe erfolgen. Ein solches Herangehen wird in einer Arbeit von Frolow und Tschenzow [2] untersucht.

Es scheint, als ob in Zukunft eine Reduktion des Problems erreicht wird, indem man die besonders schlechten (unzulässigen) Modelle in einer Klasse von Modellen wegläßt. Gegenwärtig ist es aber schwierig, diesen Standpunkt weiter zu entwickeln, weil konkrete Resultate noch fehlen.

§ 3. Einige Anwendungen

1. Die Integralgleichung des Strahlendurchgangs und deren unmittelbare Modellierung. Wir wenden uns jetzt wieder der Aufgabe des Strahlendurchgangs durch Materie zu und betrachten diese Aufgabe vom Standpunkt der Resultate, die in diesem Kapitel dargestellt worden sind. Dem im dritten Kapitel beschriebenen Modell entspricht offensichtlich ein Markoffscher Zufallsprozeß, der homogen in der Zeit ist. Dabei sind die Bewegungen der Teilchen im Phasenraum der Lage- und Geschwindigkeitskoordinaten voneinander unabhängig und identisch verteilt. Andererseits kann der Prozeß des Strahlendurchgangs durch eine Integrodifferentialgleichung (die Gleichung des Durchgangs) beschrieben werden, die ihrerseits in eine Integralgleichung übergeführt werden kann. Diese letzte wird uns hauptsächlich interessieren, weil sie eine Gegenüberstellung der unmittelbaren Modellierung mit den Lösungsmethoden für Integralgleichungen ermöglicht. Wenn man beachtet, daß die Energie des Teilchens eindeutig mit dem Betrag der Geschwindigkeit in Beziehung steht, dann kann man, wo das günstig ist, anstelle von E und $\boldsymbol{\Omega}$ auch den Geschwindigkeitsvektor $\boldsymbol{v}$ des Teilchens betrachten. Wir bezeichnen mit $u(t, \boldsymbol{r}, \boldsymbol{v})$ die Dichte der Teilchenanzahl, die im Punkt $\boldsymbol{r}$ mit der Geschwindigkeit $\boldsymbol{v}$ zur Zeit t auftritt. Wird die Funktion u so normiert, daß das Integral von u über den Phasenraum der Veränderlichen $\boldsymbol{x} = (t, \boldsymbol{r}, \boldsymbol{v})$ gleich Eins wird, dann kann u auch als Wahrscheinlichkeitsdichte der Teilchen zur Zeit t im Punkt $\boldsymbol{r}$ mit der Geschwindigkeit $\boldsymbol{v}$ gedeutet werden.

Die Integrodifferentialgleichung des Durchgangs hat für die Funktion $u(t, \boldsymbol{r}, \boldsymbol{v})$ die Gestalt (Frolow und Tschenzow [3])

$$\begin{aligned} &\frac{\partial u}{\partial t} + (\boldsymbol{v} \cdot \operatorname{grad} u) \\ &= \tilde{f}(\boldsymbol{x}) - |\boldsymbol{v}|\, \Sigma(\boldsymbol{r}, |\boldsymbol{v}|)\, u + \int_{(\boldsymbol{v})} |\boldsymbol{v}|\, \Sigma_{\mathrm{Str}}(\boldsymbol{r}, \boldsymbol{v}')\, K_1(\boldsymbol{v}', \boldsymbol{v}, \boldsymbol{r})\, u(t, \boldsymbol{r}, \boldsymbol{v}')\, d\boldsymbol{v}' \\ &\quad + \int_{(\boldsymbol{v}')} \bar{\nu}(\boldsymbol{r}, \boldsymbol{v}')\, |\boldsymbol{v}|\, \Sigma_{\mathrm{Zerf}}(\boldsymbol{r}, |\boldsymbol{v}'|)\, K_2(\boldsymbol{v}', \boldsymbol{v}, \boldsymbol{r})\, u(\boldsymbol{x})\, d\boldsymbol{v}'\,. \end{aligned} \tag{6.3.1}$$

Dabei ist $\Sigma(\boldsymbol{r}, |\boldsymbol{v}|)$ der vollständige makroskopische Wechselwirkungsquerschnitt, Σ_{Str} der makroskopische Streuquerschnitt und Σ_{Zerf} der makroskopische Zerfalls-

querschnitt; $\bar{\nu}$ ist die mittlere Anzahl der beim Zerfall gebildeten neuen Teilchen; $K_1(\boldsymbol{v}', \boldsymbol{v}, \boldsymbol{r})$ ist die Wahrscheinlichkeitsdichte dafür, daß sich bei der Streuung die Geschwindigkeit von $\boldsymbol{v}'$ auf $\boldsymbol{v}$ ändert (Streuungsindikatrix); $K_2(\boldsymbol{v}', \boldsymbol{v}, \boldsymbol{r})$ ist die Wahrscheinlichkeitsdichte dafür, daß durch den Zerfall eines Teilchens mit der Geschwindigkeit $\boldsymbol{v}'$ ein Teilchen mit der Geschwindigkeit $\boldsymbol{v}$ entsteht; $f(\boldsymbol{x})$ ist die Teilchendichte der Quelle.

Zur Vereinfachung der Darstellung werden wir nur den stationären Fall $\partial u/\partial t = 0$ mit fehlendem Zerfall, also $\Sigma_{\text{Zerf}} = 0$, ausführlich betrachten. Besonders lehrreich ist dabei die Untersuchung des schon ausführlich beschriebenen Durchgangs von γ-Quanten durch Materie. Für den auch im dritten Kapitel betrachteten Kernkaskadenprozeß kann in analoger Weise eine Stoßgleichung der Gestalt (6.3.1) formuliert werden, aber wir werden diesen Prozeß in Verbindung mit der Stoßgleichung nicht weiter betrachten. Die (6.3.1) entsprechende Integralgleichung hat die Gestalt

$$\varphi(\boldsymbol{x}) = \int K(\boldsymbol{y}, \boldsymbol{x})\, \varphi(\boldsymbol{y})\, d\boldsymbol{y} + f(\boldsymbol{x})\,. \tag{6.3.2}$$

Dabei ist $\varphi(\boldsymbol{x})$ die Dichte der Teilchenwechselwirkung mit der Materie im Punkt $\boldsymbol{x} = (\boldsymbol{r}, \boldsymbol{v})$ des Phasenraumes; $K(\boldsymbol{y}, \boldsymbol{x})$ ist die Wahrscheinlichkeitsdichte dafür, daß das Teilchen im Punkt $\boldsymbol{x}$ eine Wechselwirkung erleidet, nachdem eine Wechselwirkung im Punkt $\boldsymbol{y} = (\boldsymbol{r}', \boldsymbol{v}')$ erfolgte; $f(\boldsymbol{x})$ ist die Dichte der Wechselwirkung von Teilchen, die unmittelbar aus der Quelle kommen.

Die physikalische Bedeutung der Integralgleichung (6.3.2) ist hinreichend klar. Sie gibt an, wie sich die Wechselwirkungsdichte im Punkt $\boldsymbol{x}$ aus der durch die Quelle hervorgerufenen Dichte und aus der Wechselwirkungsdichte im ganzen Phasenraum $\mathcal{R} \times \mathcal{V}$ zusammensetzt, wobei der Übergang des Teilchens von einem beliebigen Punkt $\boldsymbol{y}$ in den Punkt $\boldsymbol{x}$ mit der Wahrscheinlichkeitsdichte $K(\boldsymbol{y}, \boldsymbol{x})$ erfolgt. Die Äquivalenz der Gleichungen (6.3.1) und (6.3.2) kann unter sehr allgemeinen Voraussetzungen über die vorkommenden Funktionen gezeigt werden. Die formale Umformung von (6.3.1) in (6.3.2) kann mit Hilfe der Laplacetransformation durchgeführt werden.

Der Kern $K(\boldsymbol{y}, \boldsymbol{x})$ der Gleichung (6.3.2) wird in der Regel substochastisch sein, d. h., es gilt

$$\int_{\mathcal{R} \times \mathcal{V}} K(\boldsymbol{y}, \boldsymbol{x})\, d\boldsymbol{x} = 1 - g(\boldsymbol{y}) \qquad \text{mit} \qquad 0 \leqq g(\boldsymbol{y}) < 1\,.$$

Die Absorptionswahrscheinlichkeit für das Teilchen beträgt also $g(\boldsymbol{y})$. Dieses Verschwinden des Teilchens kann verschiedene physikalische Ursachen haben; etwa Teilcheneinfang durch einen Kern des Stoffes oder Verlassen der Grenzen des stofferfüllten Körpers (in das Vakuum). Für die Aufgaben der Strahlenwechselwirkung stellt außerdem der Kern bei festem $\boldsymbol{y}$ die Dichte eines verallgemeinerten Maßes dar, und die Integraloperatoren wirken im allgemeinen in Räumen verallgemeinerter Funktionen. Alle Resultate von § 2 kann man auf diesen Fall erweitern, eine ausführliche Darlegung würde aber zu weit führen. Weil wir uns hier nicht speziell mit Aufgaben des Strahlentransports beschäftigen, verzichten wir auf eine Begründung dieses Problems. Für den Fall des Durchgangs von γ-Quanten durch Materie schreibt man den Kern der Integralgleichung

(6.3.2) gewöhnlich in der Form

$$K(\boldsymbol{y}, \boldsymbol{x}) = \frac{\Sigma_{\mathrm{Str}}(\boldsymbol{y})}{\Sigma(\boldsymbol{y})} \sigma_k(\boldsymbol{y} \to \boldsymbol{x}) \, \delta(\lambda - \lambda' - 1 + \boldsymbol{\Omega}\boldsymbol{\Omega}')$$
$$\times \frac{\Sigma(\boldsymbol{x}) \exp\left(-\tau(\boldsymbol{y}, \boldsymbol{x})\right)}{|\boldsymbol{r} - \boldsymbol{r}'|^2} \delta\left(\boldsymbol{\Omega} - \frac{\boldsymbol{r} - \boldsymbol{r}'}{|\boldsymbol{r} - \boldsymbol{r}'|}\right). \tag{6.3.3}$$

Dabei ist $\sigma_k(\boldsymbol{y} \to \boldsymbol{x}) \, \delta(\lambda - \lambda' - 1 + \boldsymbol{\Omega}\boldsymbol{\Omega}') = K_1(\boldsymbol{y}, \boldsymbol{x})$ die Indikatrix der Comptonstreuung; λ ist die dimensionslose Wellenlänge der γ-Quanten, die mit der in Elektronenvolt gemessenen Energie durch die Beziehung $\lambda = mc^2/E$ zusammenhängt (m Ruhemasse des Elektrons, c Lichtgeschwindigkeit). Im Fall der γ-Quanten empfiehlt es sich, anstelle der Energie E die Wellenlänge λ zu betrachten. Die Größe $\tau(\boldsymbol{y}, \boldsymbol{x})$ nennt man optische Länge zwischen den Punkten $\boldsymbol{y}$ und $\boldsymbol{x}$. Sie wird durch die Gleichung

$$\tau(\boldsymbol{y}, \boldsymbol{x}) = \int_0^{|\boldsymbol{r}-\boldsymbol{r}'|} \Sigma\,(\boldsymbol{r}' + t\boldsymbol{\Omega}, \lambda) \, dt \tag{6.3.4}$$

bestimmt. $\delta(\boldsymbol{x})$ bezeichnet die Diracsche δ-Funktion, d. h. die durch die Beziehung

$$\int_{\mathfrak{X}} f(\boldsymbol{x}) \, \delta(\boldsymbol{x} - \boldsymbol{x}_0) \, d\boldsymbol{x} = f(\boldsymbol{x}_0)$$

für alle finiten Funktionen $f(\boldsymbol{x})$ definierte verallgemeinerte Funktion. Das Auftreten der ersten δ-Funktion im Kern der Gleichung bedeutet, daß die Wellenlänge λ der γ-Quanten nach der Streuung eindeutig durch die Bewegungsrichtung nach Streuung bestimmt ist (vgl. (3.1.1)). Die zweite δ-Funktion bedeutet, daß die Integration über $\boldsymbol{\Omega}$ entartet und daß sich als Ergebnis der Wert des Integranden auf der Verbindungslinie zwischen $\boldsymbol{r}$ und $\boldsymbol{r}'$ ergibt.

Führen wir nun noch die Beziehung

$$K_2(\boldsymbol{y}, \boldsymbol{x}) = \frac{\Sigma(\boldsymbol{x}) \exp\left(-\tau(\boldsymbol{y}, \boldsymbol{x})\right)}{|\boldsymbol{r} - \boldsymbol{r}'|^2} \delta\left(\boldsymbol{\Omega} - \frac{\boldsymbol{r} - \boldsymbol{r}'}{|\boldsymbol{r} - \boldsymbol{r}'|}\right) \tag{6.3.5}$$

ein, so gilt

$$K(\boldsymbol{y}, \boldsymbol{x}) = \frac{\Sigma_{\mathrm{Str}}(\boldsymbol{y})}{\Sigma(\boldsymbol{y})} K_1(\boldsymbol{y}, \boldsymbol{x}) \, K_2(\boldsymbol{y}, \boldsymbol{x}). \tag{6.3.6}$$

Die Streuungsindikatrix $K_1(\boldsymbol{y}, \boldsymbol{x})$ hängt nicht von der Veränderlichen $\boldsymbol{r}'$ ab, und $K_2(\boldsymbol{y}, \boldsymbol{x})$ hängt nicht von den Veränderlichen λ' und $\boldsymbol{\Omega}'$ ab.

Beim Übergang zu Polarkoordinaten mit dem Ursprung in $\boldsymbol{r}'$ vereinfacht sich die Gestalt des Kernes. Günstigerweise wählt man als neue Ortskoordinaten $t = |\boldsymbol{r} - \boldsymbol{r}'|$, $\mu = \cos\vartheta = (\boldsymbol{\Omega}, \boldsymbol{\Omega}')$ und den Azimutwinkel φ. Die Funktionaldeterminante dieser Abbildung hat mit diesen Bezeichnungen den Wert $t^2 \, d\mu \, d\varphi$. Es gelten dann folgende Beziehungen (Normierungsbedingungen):

$$\int_{\mu_0}^{1} d\mu \int_0^{2\pi} d\varphi \int_0^{\lambda'} K_1(\boldsymbol{y}, \boldsymbol{x}) \, dt = 1 \qquad \text{mit} \qquad \mu_0 = \begin{cases} 1 - \lambda' & \text{für} \quad \lambda' \leqq 2\,, \\ -1 & \text{für} \quad \lambda' > 2 \end{cases}$$

(vgl. (3.1.3)) und

$$\int_0^\infty \Sigma(\boldsymbol{x}) \exp\left(-\tau(\boldsymbol{y}, \boldsymbol{x})\right) dt = 1 \tag{6.3.7}$$

(vgl. (3.1.1)). Die letzte Gleichung hat folgende Bedeutung: Wenn $\boldsymbol{r} = (x, y, z)$ einen beliebigen Wert annehmen kann (also $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < z < \infty$), dann ist $K_2(\boldsymbol{y}, \boldsymbol{x})$ für beliebige λ, μ und φ bezüglich $\boldsymbol{x}$ auf Eins normiert.

Ist also das Gebiet, in dem die Bewegung der γ-Quanten vor sich geht, der dreidimensionale euklidische Raum über der Variablen $\boldsymbol{r}$ und gilt $\Sigma(\boldsymbol{x}) > 0$ für alle $\boldsymbol{x}$, dann genügt der Kern $K(\boldsymbol{y}, \boldsymbol{x})$ der Beziehung

$$\int K(\boldsymbol{y}, \boldsymbol{x})\, d\boldsymbol{x} = \frac{\Sigma_{\text{Str}}(\boldsymbol{y})}{\Sigma(\boldsymbol{y})} = 1 - \frac{\Sigma_{\text{Abs}}(\boldsymbol{y})}{\Sigma(\boldsymbol{y})} < 1\,,$$

wobei Σ_{Abs} der Absorptionsquerschnitt der Teilchen ist. In diesem Fall kann man als Dichte $p(\boldsymbol{y} \to \boldsymbol{x})$ der Übergangswahrscheinlichkeit des Markoffschen Prozesses den Kern $K(\boldsymbol{y}, \boldsymbol{x})$ wählen. Dann ist $g(\boldsymbol{y}) = \Sigma_{\text{Abs}}(\boldsymbol{y})/\Sigma(\boldsymbol{y})$ die Wahrscheinlichkeit für die physikalische Absorption des Teilchens im Punkt $\boldsymbol{y}$. Ist das Maß der Menge jener $\boldsymbol{\Omega}$ ungleich Null, für die $\Sigma(\boldsymbol{r}' + t\boldsymbol{\Omega}, \lambda)$ mit $t \geqq a > 0$ (a fest) und beliebigen $\boldsymbol{r}'$ und λ verschwindet (das gilt beispielsweise im Fall eines räumlich beschränkten Variabilitätsbereiches von $\boldsymbol{r}$), dann kann das Teilchen das Gebiet verlassen, und die Wahrscheinlichkeit seiner Rückkehr kann Null sein. Diese Erscheinung nennt man Abfluß aus dem Gebiet. Ein solcher Zustand, bei dem das Teilchen unwiderruflich das Gebiet verläßt, stellt auch einen absorbierenden Zustand dar. Daher ist im allgemeinen Fall $g(\boldsymbol{y})$ gleich der (physikalischen) Absorptionswahrscheinlichkeit im Punkt $\boldsymbol{y}$ vermehrt um die Wahrscheinlichkeit dafür, daß sich der Punkt $\boldsymbol{y}$ außerhalb des Gebiets $\mathfrak{R}$ befindet. In beiden Fällen ist leicht zu sehen, daß die Wahl von $K(\boldsymbol{y}, \boldsymbol{x})$ als Dichte $p(\boldsymbol{y} \to \boldsymbol{x})$ zu der schon im dritten Kapitel beschriebenen unmittelbaren physikalischen Modellierung führt. Der Einfachheit halber nehmen wir weiterhin an, daß das Gebiet, in dem $\Sigma(\boldsymbol{x})$ ungleich Null ist, konvex ist. Dann ist $g(\boldsymbol{y})$ für solche $\boldsymbol{y}$, für die die räumliche Komponente von $\boldsymbol{r}$ zu $\mathfrak{R}$ gehört, gleich $\Sigma_{\text{Str}}(\boldsymbol{y})/\Sigma(\boldsymbol{y})$ und gleich Eins, wenn $\boldsymbol{r}$ außerhalb des Gebietes $\mathfrak{R}$ liegt. Man sieht leicht, daß bei der Berechnung der Austrittswahrscheinlichkeit eines Teilchens aus dem Gebiet $\mathfrak{R}$ die Beziehung $h(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{s})$ gilt, wobei $\boldsymbol{s}$ zum Rand von $\mathfrak{R}$ gehört. Wenn $\Sigma(\boldsymbol{x})$ bei hinreichend großem $|\boldsymbol{r}|$ für eine Menge $\boldsymbol{\Omega}$ positiven Maßes bei einem festen Ursprung der Polarkoordinaten verschwindet, dann fallen die Schätzungen $\hat{J}_1$ und $\hat{J}_2$ zusammen. In der Tat ist q_i in diesem Fall gleich Eins, und es gilt $\hat{J}_1 = \hat{J}_2$, wenn die Trajektorie bei Austritt des Teilchens aus dem Gebiet abgebrochen wird, und q_i ist gleich Null im entgegengesetzten Fall.

2. Einfachste Beispiele „fiktiver“ Modelle. Wählt man für $p(\boldsymbol{y} \to \boldsymbol{x})$ eine von $K(\boldsymbol{y}, \boldsymbol{x})$ verschiedene Funktion und für $\pi(\boldsymbol{x})$ eine von f verschiedene Funktion, so wird das durch die Markoffsche Kette $\{\pi, p\}$ beschriebene Modell im allgemeinen nicht dem physikalischen Bild der Strahlung entsprechen. Unter be-

stimmten Bedingungen (den Bedingungen (6.2.6)), die in diesem Fall die Absolutstetigkeit des durch $\{f, K\}$ induzierten Maßes in bezug auf das durch $\{\pi, p\}$ induzierte Maß bedeuten, können jedoch erwartungstreue Schätzungen beliebiger Mittelwerte über den Trajektorien des realen Prozesses konstruiert werden. Wir geben nun einige einfache Beispiele für die Konstruktion solcher fiktiven Modelle an.

a) Berechnung der physikalischen Absorptionswahrscheinlichkeit mit Hilfe statistischer Gewichte. Wenn uns die Austrittswahrscheinlichkeit aus einem Gebiet $\mathfrak{R}$ interessiert, dessen Ausmaße so sind, daß im Mittel mehrere Wechselwirkungen des Teilchens mit dem Stoff vor seinem Austritt aus den Grenzen von $\mathfrak{R}$ erfolgen, und wenn die Größe $\Sigma_{\text{Str}}/\Sigma$ klein ist, dann wird der Austritt eines Teilchens aus $\mathfrak{R}$ ein seltenes Ereignis sein, und die Dispersion der Schätzung kann sehr groß sein. Als Übergangswahrscheinlichkeit wählt man in diesem Fall zweckmäßigerweise

$$p(\boldsymbol{y} \to \boldsymbol{x}) = \frac{\Sigma(\boldsymbol{y})}{\Sigma_{\text{Str}}(\boldsymbol{y})} K(\boldsymbol{y}, \boldsymbol{x}) .$$

Dabei wird die Trajektorie des fiktiven Teilchens offensichtlich nicht innerhalb des Gebietes abbrechen, sondern wird immer durch Austritt aus dem Gebiet enden. Wegen (6.2.11) ist $q_{i,i+1} = \Sigma_{\text{Str}}(\boldsymbol{x}_i)/\Sigma(\boldsymbol{x}_i)$. Das Teilchen wird nicht absorbiert, aber sein „Gewicht" nimmt von Zusammenstoß zu Zusammenstoß ab. Dieses Verfahren vergrößert die Anzahl der austretenden Teilchen und verkleinert die Dispersion der Zufallsgröße, die die Austrittswahrscheinlichkeit schätzt.

b) Es sei nun $\mathfrak{R}$ eine Schicht von geringer Dicke, die Quelle liege im Innern und die Strahlung erfolgte vollständig oder größtenteils normal zur Oberfläche der Schicht. Uns interessiert der Anteil von Teilchen, die die Schicht verlassen, nachdem sie wenigstens einmal gestreut wurden. Die überwiegende Anzahl der Teilchen verläßt die Schicht ohne jede Wechselwirkung mit dem Stoff, d. h., die meisten Teilchengeschichten werden von uns nutzlos verfolgt. Die Wahrscheinlichkeit, daß ein Teilchen, das aus dem Punkt $\boldsymbol{y}$ stammt und die Bewegungsrichtung $\boldsymbol{\Omega}$ hat, die Schicht ohne Wechselwirkung verläßt, beträgt

$$q(\boldsymbol{y}) = \exp\left(-\int_0^{a(\boldsymbol{y})} \Sigma\,(\boldsymbol{y}_0 + t\boldsymbol{\Omega}, \lambda)\,dt\right).$$

Dabei ist $a(\boldsymbol{y})$ der Abstand vom Punkt $\boldsymbol{y}$ bis zur Grenze der Platte. Wir setzen

$$p(\boldsymbol{y} \to \boldsymbol{x}) = \frac{K(\boldsymbol{y}, \boldsymbol{x})}{1 - q(\boldsymbol{y})}. \tag{6.3.8}$$

Dann verläßt das Teilchen niemals die Schicht, und der uns interessierende Anteil der Teilchen, die wenigstens einmal gestreut werden, kann genauer geschätzt werden. Ein Abbrechen der Trajektorie ist nur möglich als Folge einer physikalischen Absorption. Das Gewicht des Teilchens wird durch die Größen $q_{i,i+1} = 1 - q(\boldsymbol{x}_i)$ bestimmt. Die Weglänge des Teilchens wird nicht so wie bei der direkten Modellierung berechnet. In Übereinstimmung mit (6.3.8) ist nämlich die Weg-

länge l_i von $\boldsymbol{r}_i$ bis $\boldsymbol{r}_{i+1}$ mit der Dichte

$$\Sigma(\boldsymbol{r}_i + l_i\boldsymbol{\Omega}_i, \lambda_i) \exp\left(-\int_0^{l_i} \Sigma(\boldsymbol{r}_i + t\boldsymbol{\Omega}_i, \lambda_i)\,dt\right)$$
$$\times\left(1 - \exp\left(-\int_0^{a(x_i)} \Sigma(\boldsymbol{r}_i + t\boldsymbol{\Omega}_i, \lambda_i)\,dt\right)\right)^{-1}$$

verteilt. Wenn Σ innerhalb der Schicht nicht von $\boldsymbol{r}$ abhängt, erhalten wir für l_i den Ausdruck

$$l_i = \frac{1}{\Sigma}\log\left[1 - \alpha\big(1 - \exp\left(-\Sigma a(\boldsymbol{x}_i)\right)\big)\right]. \tag{6.3.9}$$

Wie leicht zu sehen ist, vergrößert die beschriebene Methode den Rechenaufwand bei der Modellierung der Trajektorien, weil die Berechnung des Abstandes bis zur Grenze in jedem Schritt und die Berechnung der in (6.3.9) und in den $Q_i^{(k)}$ eingehenden Exponentialfunktion erforderlich ist. Bei einer wirklichen Rechnung muß man diesen Umstand beachten. Besonders aufwendig wird die Rechnung, wenn Σ von $\boldsymbol{r}$ abhängt und als Gebiet $\mathcal{R}$ keine Platte, sondern ein Gebiet komplizierterer Form gewählt wird.

Wählt man für die Dichte $p(\boldsymbol{y} \to \boldsymbol{x})$ die Funkion

$$\frac{K(\boldsymbol{y}, \boldsymbol{x})\,\Sigma(\boldsymbol{y})}{\Sigma_{\mathrm{Str}}(\boldsymbol{y})\,\big(1 - q(\boldsymbol{y})\big)},$$

so ist $g(\boldsymbol{y}) = 0$, alle Trajektorien sind unendlich, und man muß sie solange verfolgen, bis die Größe $Q_i^{(k)}$ vernachlässigbar klein wird.

Zur Dispersionsverkleinerung bei Schätzungen von Funktionalen in Aufgaben der behandelten Art werden oft Informationen über die Lösung der adjungierten Gleichung benutzt. Die Grundlage dafür bilden die allgemeinen Überlegungen, die im vorigen Paragraphen dargestellt wurden. In einigen Fällen gelingt es, für einfachere Aufgaben die adjungierte Gleichung analytisch zu lösen, beispielsweise in einfachen Gebieten unter der Voraussetzung konstanter Wirkungsquerschnitte und anderer vereinfachender Annahmen. Man kann versuchen, die auf diese Weise gewonnene Lösung zur Dispersionsverkleinerung zu benutzen. Der Effekt wird gewöhnlich experimentell durch Schätzung der Dispersion auf Grund einer Stichprobe mäßigen Umfangs überprüft.

Man kann aber auch wie in Kap. IV, § 5, von einem anderen Standpunkt ausgehen und eine positive von m Parametern abhängige Funktion $\vartheta(\boldsymbol{x}, a_1, \ldots, a_m)$ vorgeben. Die Dichte $p(\boldsymbol{y} \to \boldsymbol{x})$ der Übergangswahrscheinlichkeit wird dann in der Form

$$p(\boldsymbol{y} \to \boldsymbol{x}) = \frac{K(\boldsymbol{y}, \boldsymbol{x})\,\vartheta(\boldsymbol{y}, a_1, \ldots, a_m)}{\vartheta(\boldsymbol{x}, a_1, \ldots, a_m)}$$

gewählt. Dabei muß offensichtlich die Ungleichung

$$\int \frac{K(\boldsymbol{y}, \boldsymbol{x})}{\vartheta(\boldsymbol{x}, a_1, \ldots, a_m)}\,d\boldsymbol{x} \leqq \frac{1}{\vartheta(\boldsymbol{y}, a_1, \ldots, a_m)}$$

für alle Werte $a_1, \ldots, a_m$ einer gewissen Menge erfüllt sein. Bei einer komplizierten Gestalt des Kernes $K(\boldsymbol{y}, \boldsymbol{x})$ kann diese Forderung schwerwiegend sein. Für die Schätzung $\hat{J}_1$ ist bekannt, daß die Dispersion für eine Übergangsdichte geeigneter Gestalt verschwindet. Daher kann man die Aufgabe nach einer solchen Wahl der Parameter $a_1, \ldots, a_m$ stellen, daß die Dispersion von $\hat{J}_1$ minimal wird. Ein Erfolg hängt hier viel von der geschickten Wahl der Funktion ϑ ab. Die Größen $a_1, \ldots, a_m$ kann man auch durch das Bestimmen einer Näherungslösung der adjungierten Gleichung in der Gestalt $\vartheta(\boldsymbol{x}, a_1, \ldots, a_m)$ schätzen. Dazu ist es sinnvoll, Variationsmethoden zu benutzen. Freilich gibt es noch andere Verfahren, Parameter in die Dichte (und damit in die Größen $Q_i^{(k)}$) einzuführen. Eine weitverbreitete Methode stellt die Exponentialtransformation dar (Fano u. a. [1], Michailow [3]), auf die wir nicht näher eingehen. Besonders wichtig ist es, Methoden der Dispersionsverkleinerung bei der Berechnung kleiner Wahrscheinlichkeiten mit der Monte-Carlo-Methode auszuarbeiten, beispielsweise die Wahrscheinlichkeit, daß γ-Quanten in ein kleines Gebiet des Phasenraumes fallen, oder die Wahrscheinlichkeit, daß eine dicke Schicht durchstrahlt wird, wenn die Strahlung um eine Größenordnung von 10^{-4} abgeschwächt wird. In diesem letzten Fall erreicht bei unmittelbarer Modellierung von 10^4 Teilchen im Mittel nur eines die Grenze des Gebietes, und der Rechenaufwand zur Erzielung eines unmittelbaren Resultates wird außerordentlich groß.

Als ein Grenzfall ergibt sich der Fall, daß die Strahlung in einem Punkt des Raumes $\mathcal{R}$ zu berechnen ist. Die Wahrscheinlichkeit, daß ein Teilchen einen bestimmten Punkt des Raumes berührt, ist offensichtlich gleich Null, und die Methoden der unmittelbaren Modellierung werden unwirksam. Es gibt aber spezielle Verfahren, die eine Lösung auch dieser Aufgabe ermöglichen.

3. Lokale Methoden. Das anschließend zu beschreibende Verfahren kann bei der Lösung beliebiger Integralgleichungen Anwendung finden. Für die Gleichung des Strahlentransports besitzt es allerdings gewisse Besonderheiten, die mit der konkreten Gestalt des Kernes zusammenhängen. Das Verfahren arbeitet folgendermaßen. Aus (6.2.1) folgt

$$\varphi(\boldsymbol{x}^0) = \int K(\boldsymbol{y}, \boldsymbol{x}^0)\, \varphi(\boldsymbol{y})\, d\boldsymbol{y} + f(\boldsymbol{x}^0)$$

mit einem festen Punkt $\boldsymbol{x}^0$ aus $\mathcal{R} \times \mathcal{V}$. Zur Berechnung der Größe $\int K(\boldsymbol{y}, \boldsymbol{x}^0)\, \varphi(\boldsymbol{y})\, d\boldsymbol{y}$ kann eine beliebige erwartungstreue Schätzung (z. B. $\hat{J}_1$ oder $\hat{J}_2$) benutzt werden, wenn man $h(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}^0)$ setzt. Die Berechnung von $f(\boldsymbol{x}^0)$ bereitet keine Schwierigkeiten. So kann man den Wert der Lösung der Integralgleichung in einem festen Punkt $\boldsymbol{x}^0$ schätzen. Für die Gleichung des Strahlentransports ist dazu die Wechselwirkungsdichte mit dem Stoff im Punkt $\boldsymbol{x}^0$ zu schätzen. Die Strahlung, die als Lösung der Integrodifferentialgleichung erscheint, ist in einfacher Weise mit $\varphi(\boldsymbol{x}^0)$ verknüpft. (Bei der Berechnung der Strahlung muß man $h(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}^0)/\Sigma_{\mathrm{Str}}(\boldsymbol{x})$ wählen.) Leider hat der Kern $K(\boldsymbol{y}, \boldsymbol{x})$ der Strahlentransportgleichung starke Singularitäten, wie etwa das Auftreten der δ-Funktion oder der Größe $|\boldsymbol{r} - \boldsymbol{r}'|^2$ im Nenner. Die letzte Singularität führt dazu, daß die Dispersion der Schätzungen $\hat{J}_1, \hat{J}_2$ und anderer Schätzungen unendlich wird, wenn der Punkt $\boldsymbol{r}^0$ innerhalb des Gebietes von $\mathcal{R}$ liegt, wo $\Sigma > 0$ ist. Spezielle Ver-

fahren, die die Herleitung von Schätzungen mit endlicher Dispersion ermöglichen, sind in numerischer Hinsicht recht aufwendig (ein solches Verfahren beschreiben wir anschließend). Daher benutzt man für eine grobe Schätzung der Strahlung manchmal die Schätzungen $\hat{J}_1$ und $\hat{J}_2$ mit $h(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}^0)/\Sigma_{\text{Str}}(\boldsymbol{x})$, d. h. das Schema B entsprechend der Einteilung von Kap. II, § 3. Die Konvergenzordnung beträgt dabei $O(N^{-1/3})$, und es gelten alle Überlegungen, die früher im Zusammenhang mit dem Schema B gemacht wurden.

Eines der Verfahren, das die Herleitung einer Schätzung mit endlicher Dispersion ermöglicht (Kalos [1]), besteht in folgendem: Zur Schätzung der Größe $\int K(\boldsymbol{x}, \boldsymbol{x}^0)\, \varphi(\boldsymbol{x})/\Sigma_{\text{Str}}(\boldsymbol{x})\, d\boldsymbol{x}$ wird eine spezielle Verteilung des Punktes $\boldsymbol{x}$ gewählt (die Singularität wird in der Dichte berücksichtigt). So kann für die Berechnung des Beitrages beim Übergang von $\boldsymbol{x}_0$ zu $\boldsymbol{x}_1$ die Verteilungsdichte

$$\tilde{U}(\boldsymbol{x}_0 \to \boldsymbol{x}_1) = U(\boldsymbol{r}_1) = C \frac{|\boldsymbol{r}_0 - \boldsymbol{r}^{(0)}|}{|\boldsymbol{r}_1 - \boldsymbol{r}_0|^2\, |\boldsymbol{r}_1 - \boldsymbol{r}^{(0)}|^2}.$$

gewählt werden. Analog wird im allgemeinen Fall die Trajektorie $\boldsymbol{x}_0 \to \boldsymbol{x}_1 \to \cdots \to \boldsymbol{x}_k$ verfolgt, die einer Wahrscheinlichkeitsdichte

$$\pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \cdots p(\boldsymbol{x}_{k-2} \to \boldsymbol{x}_{k-1})\, U(\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k)$$

entspricht. Obwohl die Kette nicht homogen ist, prüft man leicht, daß mit

$$q_i = \frac{f(\boldsymbol{x}_0)\, K(\boldsymbol{x}_0, \boldsymbol{x}_1)}{\pi(\boldsymbol{x}_0)\, p(\boldsymbol{x}_0 \to \boldsymbol{x}_1)} \cdots \frac{K(\boldsymbol{x}_{i-2}, \boldsymbol{x}_{i-1})}{p(\boldsymbol{x}_{i-2} \to \boldsymbol{x}_{i-1})} \cdot \frac{K(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i)}{\tilde{U}(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i)}$$

die Schätzungen $\hat{J}_1$ und $\hat{J}_2$ erwartungstreu sind. Dabei enthält q_i den Faktor $|\boldsymbol{r}_i - \boldsymbol{r}^0|^2$, und die Schätzungen $\hat{J}_1$ und $\hat{J}_2$ für $\int K(\boldsymbol{y}, \boldsymbol{x}^0)\, \varphi(\boldsymbol{y})/\Sigma_{\text{Str}}(\boldsymbol{y})\, d\boldsymbol{y}$ enthalten im Nenner nicht mehr die Größe $|\boldsymbol{r}_i - \boldsymbol{r}^0|^2$. Jedoch enthält der Nenner die Größe $|\boldsymbol{r}_{i-1} - \boldsymbol{r}^0|$ (aus dem Zähler von U), aber das zweite Moment der Schätzung ist schon endlich. Es bleibt zu bemerken, daß das dritte und die höheren Momente der Schätzung nicht existieren, so daß die Konvergenz der Verteilung der Schätzung gegen die Normalverteilung sehr langsam erfolgt. Die gewöhnliche Schätzung der Dispersion (1.3.5) ist recht ungenau, und die Herleitung eines Konfidenzintervalls auf Grund der Normalverteilung ist nicht gerechtfertigt. Man kann kompliziertere Verfahren angeben, die die Endlichkeit des dritten und der höheren Momente der Verteilung der Schätzung sichern. Möglich ist aber auch die Wahl einer anderen Dichte U. Eine Untersuchung dieser Probleme findet man in der Spezialliteratur (z. B. Kalos [1], Ermakow u. a. [1], Chissamutdinow [1]. Ist der Teilchenstrom in einer hinreichend großen Umgebung von $\boldsymbol{r}^0$ durch eine glatte Funktion darstellbar, kann man folgendermaßen vorgehen. Wie im fünften Kapitel beschrieben, werden die Fourierkoeffizienten des Teilchenstromes als Funktion von $\boldsymbol{r}^0$ berechnet. Dann wird der Wert der Fourierreihenteilsumme für den Teilchenstrom im Punkt $\boldsymbol{r}^0$ berechnet, und dieser Wert anstelle des gesuchten benutzt. Diese Methode benötigt eine strenge theoretische Begründung, obwohl auch die empirische Schätzung der Fehler des Resultats möglich ist. Der Vorteil der beschriebenen Berechnungsmethode ist offensichtlich, weil die Änderungen im Modellierungsprogramm der Trajektorien unbedeutend sind. Es besteht nur ein größerer Speicherbedarf für das Auflaufen der Summen, mit denen die Fourierkoeffizienten geschätzt werden.

Dieser letzte Umstand ist auch nur dann wesentlich, wenn die Energie- oder Winkelverteilung des Teilchenstromes in einem Punkt in Gestalt eines Histogramms oder durch die Fourierkoeffizienten nach einem Funktionensystem berechnet werden. Der zusätzliche Speicherbedarf kann dann beträchtlich sein.

Schließlich kann man zur Bestimmung des Teilchenstromes in einem Punkt die adjungierte Gleichung benutzen. Das allgemeine Vorgehen im Fall der adjungierten Gleichung bleibt natürlich gegenüber dem bisherigen ungeändert, obwohl es einige besondere Details in numerischer Hinsicht gibt. Vor allem ist folgendes zu bemerken: Spielt $f(\boldsymbol{x})$ in Gleichung (6.3.1) die Rolle einer Strahlenquelle, so spielt $h(\boldsymbol{x})$ die Rolle eines Empfängers, der die ankommenden Teilchen registriert. Bei der Lösung der adjungierten Gleichung vertauschen sich die Rollen von $f(\boldsymbol{x})$ und $h(\boldsymbol{x})$ (es wird angenommen, daß $h(\boldsymbol{x})$ nichtnegativ und normiert ist). Die Wahrscheinlichkeit, daß ein Teilchen im Punkt $\boldsymbol{x}$ erzeugt wird, ist mit der Dichte $h(\boldsymbol{x})$ verteilt. Nachdem das Teilchen mit gewissen Übergangswahrscheinlichkeiten in neue Punkte übergeht, wird es in Empfängern registriert, die mit der Dichte $f(\boldsymbol{x})$ verteilt sind. Die direkte (primale) Aufgabe mit punktförmiger Quelle aber ausgedehntem Empfänger ($h(\boldsymbol{x})$ ist in einem größeren Gebiet des Phasenraumes ungleich Null) bietet bei der Lösung keine besonderen Schwierigkeiten. Ist aber der Empfänger in einem Punkt konzentriert, dann gibt es Schwierigkeiten, über die anschließend zu sprechen sein wird. Für die adjungierte Gleichung gilt entgegengesetztes Verhalten. Daher ist es zweckmäßig, bei Aufgaben, in denen $f(\boldsymbol{x})$ in einem größeren Gebiet nicht verschwindet, aber $h(\boldsymbol{x})$ nur in einem kleinen Gebiet des Phasenraumes ebenfalls nicht verschwindet, zur Berechnung des Funktionals $(h, \varphi) = (f, \varphi^*)$ die adjungierte Gleichung zu lösen. Bestimmte Schwierigkeiten ergeben sich dabei aus folgendem Grund. Der Kern $K(\boldsymbol{x}, \boldsymbol{y})$ hat hier keinen unmittelbaren physikalischen Sinn, und das Integral $\int K(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}$ kann größer als Eins sein. Eine ausführliche Beschreibung der Lösungsverfahren für die adjungierte Gleichung findet man in den Arbeiten von Kalos [2] und Chissamutdinow [2].

4. Eine allgemeine Bemerkung zur Lösung der Stoßgleichungen. Der Durchgang von γ-Quanten durch Materie bildet ein verhältnismäßig einfaches aber typisches Beispiel für diesen Aufgabentyp. Die Aufgabe für die Neutronenstrahlung ist in rechnerischer Hinsicht komplizierter (falls nicht wesentliche Vereinfachungen des physikalischen Sachverhalts vorgenommen werden). Aber fast alle numerischen Verfahren, die im Fall der γ-Quanten benutzt werden, können auch im Fall der Neutronen Anwendung finden. Die vorangegangene Beschreibung der grundlegenden Besonderheiten bei Aufgaben der Teilchenstrahlung ist verhältnismäßig einfach gehalten und sollte keine Illusionen hervorrufen. Die realen Aufgaben sind sowohl in bezug auf die Lage der räumlichen Gebiete zueinander als auch in bezug auf die zu berücksichtigenden physikalischen Effekte recht kompliziert. Die Auswahl eines Modells erfordert bei realen Aufgaben ernsthafte theoretische Untersuchungen und vorbereitende Experimente an kleinen Stichproben. Aufgaben der Reaktorberechnung wurden von uns nicht berührt. Dort spielt der Teilchenzerfall eine wesentliche Rolle, und man muß die Eigenschaften der stationären Verteilung des Markoffschen Prozesses bestimmen. Eine ausführliche Darstellung aller dieser Fragen würde wesentlich mehr Kenntnisse in der Reaktorphysik erfordern, als sie in diesem Buch enthalten sind. Außerhalb unserer Betrachtung bleiben auch

viele Methoden der Dispersionsverkleinerung, die in allgemeiner Form im vorigen Paragraphen dargestellt wurden, deren konkrete Anwendung auf die Gleichungen der Teilchenstrahlung aber nicht immer trivial ist. Wir kommen nun zur Darstellung eines Beispiels für die Modellierung einer Markoffschen Kette zur Lösung von Aufgaben der statistischen Physik.

5. Modellierung einer Markoffschen Kette in Zusammenhang mit Aufgaben der statistischen Physik. Es wird ein System von N einatomigen Teilchen betrachtet. Wir bezeichnen mit $\boldsymbol{r}_i$ die Lage des Schwerpunkts und mit $\boldsymbol{v}_i$ die Geschwindigkeit des i-ten Teilchens, $i = 1, \ldots, N$. Eine Menge $\boldsymbol{r}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_N$ nennen wir *Zustand des Systems*. Die Hamiltonfunktion $H(\boldsymbol{r}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_N)$ ist die Summe von kinetischer und potentieller Energie:

$$H(\boldsymbol{r}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_N) = \sum_{i=1}^{N} \frac{mv_i^2}{2} + U_N(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$$

(m Masse des Teilchens). Wir machen nun die vereinfachende Annahme, daß die potentielle Energie durch die paarweise Wechselwirkung bestimmt wird:

$$U_N(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \sum_{i<j} \Phi(|\boldsymbol{r}_i - \boldsymbol{r}_j|)\,,$$

wobei Φ eine gegebene Funktion ist. Aus der statistischen Physik ist bekannt, daß bei genügend großem N für die Wahrscheinlichkeitsdichte des Zustandes $\boldsymbol{r}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_N$

$$a_N^{-1} \exp\left\{-\frac{1}{kT} H(\boldsymbol{r}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_N)\right\} \tag{6.3.10}$$

gilt. Dabei ist a_N eine Normierungskonstante, k eine Konstante, und T kann als Temperatur gedeutet werden. Weil $\boldsymbol{v}_i$ nur in den Ausdruck für die kinetische Energie eingeht und dieser verhältnismäßig einfach ist, kann man (6.3.10) nach allen $\boldsymbol{v}_i$ analytisch integrieren und erhält so die Dichte für die Aufenthaltswahrscheinlichkeit in den Punkten $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ in der Form

$$b_N^{-1} \exp\left\{-\frac{1}{kT}\; V(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)\right\}. \tag{6.3.11}$$

Dabei ist b_N (das sogenannte Lageintegral) eine Normierungskonstante. Die Größe a_N kann durch b_N ausgedrückt werden. Verschiedene thermodynamische Größen wie Druck, Entropie u. a. können durch die partiellen Ableitungen von a_N ausgedrückt werden, d. h. durch Mittelwerte der Gestalt

$$\overline{F} = \frac{\int F(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) \exp\{-V/kT\}\, d\boldsymbol{r}_1 \cdots d\boldsymbol{r}_N}{\int \exp\{-V/kT\}\, d\boldsymbol{r}_1 \cdots d\boldsymbol{r}_N}.$$

Unsere Aufgabe besteht daher in der Berechnung von $\overline{F}$ bei einem gegebenen F. Wir zerlegen dazu den ganzen Raum in Würfel mit hinreichend großer Seitenlänge

und werden annehmen, daß die Aufgabe in bezug auf die Seitenflächen des Würfels symmetrisch ist (wenn das Teilchen irgendeine Fläche schneidet und aus dem Würfel austritt, taucht es automatisch auf der entgegengesetzten Seite auf). Innerhalb des Würfels wird nun ein Netz mit kleiner Schrittweite konstruiert, und es wird angenommen, daß sich das Teilchen nur in den Knoten des Netzes befinden kann. Alle Knoten werden der Reihe nach durchnumeriert $j = 1, \ldots, n$, und man kann nun annehmen, daß

$$b_N \approx \tilde{b}_N = \sum_{j=1}^{n} \exp\left\{-\frac{V_j}{kT}\right\} \qquad \text{und} \qquad \overline{F} \approx \sum_{j=1}^{n} F_j \exp\left\{\frac{V_j}{kT}\right\}$$

gilt. Nun bildet man eine Markoffsche Kette mit den stationären Wahrscheinlichkeiten

$$g_j = b_N^{-1} \exp\left\{-\frac{V_j}{kT}\right\}.$$

Dazu muß man als Elemente der stochastischen Matrix solche $p_{j,l}$ wählen, die die Bedingung

$$\sum_{j=1}^{n} p_{j,l} g_j = g_l \tag{6.3.12}$$

erfüllen, d. h., es verlassen genau so viele Teilchen den Punkt mit dem Index j wie dann von anderen Punkten in den Punkt j kommen. Die Aufgabe ist insbesondere dann gelöst, wenn die Gleichgewichtsbedingung

$$g_j p_{j,l} = g_l p_{l,j} \tag{6.3.13}$$

gilt. Wenn man (6.3.13) über j summiert, dann erhält man offensichtlich (6.3.12).

Gilt daher

$$p_{j,l} \exp\left\{\frac{V_l}{kT}\right\} = p_{l,j} \exp\left\{\frac{V_j}{kT}\right\},$$

so ist (6.3.12) erfüllt. Weiter verfährt man nun so: Man wählt einen Punkt des Netzes mit dem Index l und bringt dort ein Teilchen an. Mit der gleichen Wahrscheinlichkeit wählt man einen anderen Punkt mit dem Index j und berechnet die ihm entsprechende potentielle Energie. Ist $U_l \leqq U_j$, so vollzieht sich ein vollständiger Übergang von j nach l. Gilt aber $U_l > U_j$, dann vollzieht sich der Übergang in den l-ten Punkt mit der Wahrscheinlichkeit $\exp\{-(U_l - U_j)/kT\}$, d. h. falls $\alpha > \exp\{-(U_l - U_j)/kT\}$ für den Stichprobenwert α gilt. Mit der Komplementärwahrscheinlichkeit findet kein Übergang statt, d. h., das Teilchen bleibt am Ort. Wegen der Symmetrie eines solchen Verfahrens (alle Teilchen werden der Reihe nach so behandelt) ist (6.3.12) erfüllt. Das arithmetische Mittel von F über alle Lagen liefert die gewünschte Schätzung. Die Parameter der zufälligen Verschiebung müssen entsprechend den physikalischen Eigenschaften des Stoffes gewählt werden, klein für Kristalle, größer für Flüssigkeiten und noch größer für Gase. Ausführlich werden diese Fragen von FISHER [1] untersucht.

§ 4. Bemerkungen über die Lösung anderer Aufgaben der numerischen Mathematik

1. Differentialgleichungen. Die prinzipielle Möglichkeit zur Lösung von Differentialgleichungen mit der Monte-Carlo-Methode ergibt sich aus den Resultaten von § 1. Nachdem man ein Netz mit genügend kleiner Schrittweite konstruiert hat und so die Differentialgleichung mit hinreichender Genauigkeit durch ein lineares algebraisches Gleichungssystem ersetzt hat, ist es leicht, die Modellierung durch eine Markoffsche Kette anzuwenden. Aber in diesem wie auch in vielen anderen Fällen bedeutet eine einfache Herleitung des stochastischen Modells noch nicht seine Effektivität in numerischer Hinsicht. Das Verfolgen der Irrfahrt auf einem Netz mit kleiner Schrittweite kann aufwendiger sein als die Lösung des Gleichungssystems mit anderen Methoden (z. B. durch sukzessive Approximation). Daher ist es wichtig, verschiedene andere stochastische Modelle für Differentialgleichungen zu untersuchen. Es gibt ziemlich viele Modelle solcher Art (z. B. Müller [1], Djadkin und Shukowa [1], Dynkin und Juschkewitsch [1]). In den meisten Fällen werden die Differentialgleichungen zur Lösung theoretischer Probleme benutzt, die mit zufälligen Prozessen verknüpft sind, aber seltener sind numerische Gesichtspunkte ausgearbeitet. Wir beschränken uns daher nur auf einige Bemerkungen über den Zusammenhang zwischen Differentialgleichungen und stochastischen Modellen.

Vor allem ist zu bemerken, daß man die Lösung des Cauchyproblems für ein System gewöhnlicher linearer Differentialgleichungen durch Modellierung eines Bedienungssystems erhalten kann. Gut bekannt ist beispielsweise, daß ein Poissonsches System mit exponentialverteilter Bedienungszeit durch ein gewöhnliches Differentialgleichungssystem mit einer im allgemeinen geraden Anzahl von Unbekannten beschrieben wird. Ein geschlossenes System, d. h. ein solches, bei dem die ausgehenden Forderungen von neuem in das System eintreten, wird durch ein System mit m Unbekannten beschrieben, wenn die Anzahl der Forderungen im System gleich m ist. Die einfachste derartige Differentialgleichung

$$y'(t) = -\lambda(t)\, y(t)\,, \qquad t \geqq 0\,,$$

ist mit einem Poissonschen Strom der Intensität $\lambda(t)$ verknüpft. Ist $y(0) = 1$, so bedeutet $y(t)$ die Wahrscheinlichkeit, daß bis zum Zeitpunkt t kein Ereignis eingetreten ist. Übrigens ist die Lösung eines gewöhnlichen linearen Differentialgleichungssystems mit nicht zu vielen Unbekannten in der Regel kein kompliziertes numerisches Problem, so daß keine Notwendigkeit für die Anwendung der Monte-Carlo-Methode besteht.

Randwertaufgaben sind vom numerischen Standpunkt aus wesentlich interessanter. Da sich Randwertaufgaben linearer Differentialgleichungen in Fredholmsche oder Volterrasche Integralgleichungen zweiter Art überführen lassen, kann auf § 3 verwiesen werden. Die Darstellung der Potentialtheorie und der Beweis der Äquivalenz von Randwertaufgaben für partielle Differentialgleichungen mit den entsprechenden Integralgleichungen findet sich in vielen Lehrbüchern (z. B. Wladimirow [1]). Es ist aber zu bemerken, daß die Konstruktion der den Kern bestimmenden Greenschen Funktion ein kompliziertes Problem sein kann, insbe-

sondere dann, wenn Gleichungen mit veränderlichen Koeffizienten betrachtet werden. Daher besteht ein bedeutendes Interesse an anderen Verfahren, die unmittelbar die stochastischen Modelle mit den Differentialgleichungen verbinden, sogar wenn sie zu analogen Resultaten führen. Sehr interessant ist in dieser Beziehung ein Modell (KAC [2]), das den Poissonschen Prozeß mit der Wellengleichung verbindet.

Erfolgreiche Berechnungen wurden hauptsächlich für die Poissonsche Gleichung, die Schrödinger-Gleichung und die Wärmeleitungsgleichung durchgeführt (vgl. z. B. HAJI-SHEIKH und SPARROW [1], DJADKIN und SHUKOWA [1], ELEPOW und MICHAILOW [1]). Eng verknüpft mit diesen Gleichungen sind die Integration nach einem Wienerschen Maß und der Prozeß der Brownschen Bewegung. Dieser Zusammenhang ist für die Wärmeleitungsgleichung leicht zu zeigen, wenn man davon ausgeht, daß die Fundamentallösung der Wärmeleitungsgleichung mit der Dichte der Normalverteilung übereinstimmt. Entsprechende Resultate für die Schrödinger-Gleichung findet man in den Arbeiten von KAC [1], DJAKIN und SHUKOWA [1] und für die Poissonsche Gleichung bei DYNKIN und JUSCHKEWITSCH [1].

Der unmittelbare Übergang vom Differenzenschema für die Lösung einer partiellen Differentialgleichung zu einem Integral über einem Maß in einem Funktionenraum wird in einer Arbeit von W. J. KRYLOW [1] betrachtet.

2. Die Lösung nichtlinearer Probleme mit der Monte-Carlo-Methode wird gewöhnlich auf die Lösung einer Folge linearer Aufgaben zurückgeführt. So löst man beispielsweise die nichtlineare Integralgleichung

$$\varphi(\boldsymbol{x}) = \int K(\boldsymbol{x}, \boldsymbol{y}, \varphi(\boldsymbol{x}))\, \varphi(\boldsymbol{y})\, d\boldsymbol{y} + f(\boldsymbol{x}) \tag{6.4.1}$$

durch sukzessive Approximation nach dem Schema

$$\varphi^{(k)}(\boldsymbol{x}) = \int K(\boldsymbol{x}, \boldsymbol{y}, \varphi^{(k-1)}(\boldsymbol{y}))\, \varphi^{(k)}(\boldsymbol{y})\, d\boldsymbol{y} + f(\boldsymbol{x})\,, \qquad k = 1, 2, \ldots \tag{6.4.2}$$

Für jedes feste k ist die Gleichung (6.4.2) linear, und sie kann mit einer der früher beschriebenen Methoden gelöst werden. Die Konvergenz dieser sukzessiven Approximation ebenso wie die Konvergenz der Neumannschen Reihen für jedes k werden dabei vorausgesetzt.

Zu einer Gleichung oder einem System von Gleichungen der Art (6.4.1) kommt man beispielsweise bei Aufgaben der Gasdynamik (WALLANDER [1]). Auf diesem Gebiet wurde eine große Zahl numerischer Experimente durchgeführt, die zu bemerkenswerten Ergebnissen führten (z. B. BOGATSCHEWA u. a. [1], HAVILAND u. a. [1]). Obwohl die Linearisierung (6.4.2) der Gleichung (6.4.1) nicht die einzige und auch nicht immer die effektivste Möglichkeit ist, führen doch physikalische Überlegungen dorthin. Vom physikalischen Standpunkt aus bedeutet die Nichtlinearität der Gleichung (6.4.1) das Vorhandensein einer Wechselwirkung zwischen den Gasmolekülen (wenn man eine Aufgabe der Gasdynamik betrachtet). Bei der Modellierung der Molekülbewegung wird zuerst irgendeine Anfangsverteilung ihrer Phasenkoordinaten, also $\varphi^{(0)}(\boldsymbol{x})$ vorgegeben. In dem durch diese Anfangsverteilung gebildeten Feld verfolgt man nun die Bewegung einer Gruppe von Molekülen und bringt dann Korrekturen an der Anfangsverteilung an, was im wesentlichen der Benutzung von Formel (6.4.2) entspricht. Es wäre interessant festzustellen, ob auch eine inhaltliche physikalische Deutung der anderen Linearisierungsmethoden

möglich ist, z. B. der Newtonschen Methode zur Lösung nichtlinearer Funktionalgleichungen.

Die Lösung nichtlinearer Gleichungen ist in numerischer Hinsicht eine recht schwierige Aufgabe. Vor allem ist es notwendig, alle Werte der gesuchten Funktion sogar in dem Fall zu ermitteln, in dem das Schlußresultat nur aus einer Zahl, einem gewissen Funktional der Lösung φ, besteht. Ist die Aufgabe mehrdimensional, so erfordert das viel zusätzlichen Speicherraum der EDVA. Eine zweite Frage, die bei der Lösung nichtlinearer Aufgaben mit der beschriebenen Methode entsteht, betrifft die Fehleranalyse, da es hier wesentlich mehr Fehlerquellen gibt, nämlich sowohl stochastische als auch deterministische. Eine recht ausführliche Fehleranalyse für die Lösung der nichtlinearen Boltzmanngleichung findet man in der Arbeit von HICKS und SMITH [1]. Insgesamt stehen die Untersuchungen zum Problem der Lösung nichtlinearer Aufgaben mit der Monte-Carlo-Methode allerdings erst am Anfang. Es bleibt zu bemerken, daß in einigen Arbeiten über die Lösung nichtlinearer Aufgaben die klassischen Methoden der numerischen Analysis mit der Monte-Carlo-Methode verknüpft werden. So wurde eine Differenzenmethode in Verbindung mit der Monte-Carlo-Methode zur Lösung der Gleichung des Strahlentransports bei Berücksichtigung der Temperatureffekte von FLECK [1] benutzt. In der Arbeit von ANOLIK und BARANZEW [1] wurden mit der Monte-Carlo-Methode Integrale berechnet, die bei der analytischen Lösung von Aufgaben der Gasdynamik entstehen.

In Zusammenhang mit der Lösung nichtlinearer Aufgaben muß auch die Arbeit von TSUDO und KIYONO [1] genannt werden, in der ein physikalisches Modell zur Lösung nichtlinearer algebraischer Gleichungssysteme betrachtet wird. Obwohl in dieser Arbeit die dem Modell zugrunde liegende Idee weniger ausführlich als die im einzelnen streng begründete Methode betrachtet wird, ist es angebracht, diese Idee darzustellen, weil sie von prinzipiellem Interesse ist und verhältnismäßig weit auseinanderliegende Gebiete der Mathematik verbindet. Es sei ein System nichtlinearer Gleichungen der Gestalt

$$F_i(x_1, \ldots, x_n) = 0\,, \qquad i = 1, \ldots, n\,, \tag{6.4.3}$$

gegeben, wobei die $F_i(x_1, \ldots, x_n)$ nach jeder Veränderlichen x_j differenzierbar sind. Wir nehmen an, daß das System eine endliche Anzahl reeller Lösungen besitzt, die es zu bestimmen gilt. Die darzustellende Methode gibt nicht die Möglichkeit, die Vielfachheit einer Lösung des Systems festzustellen, und wir werden für das weitere alle Lösungen als einfach voraussetzen. Wir ordnen dem System (6.4.3) die Funktion

$$U(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i^2 F_i^2(x_1, \ldots, x_n) \tag{6.4.4}$$

mit von Null verschiedenen reellen Konstanten a_i zu. Das nun zu bestrachtende Modell steht in Beziehung zu Aufgaben der statistischen Physik (vgl. § 3 dieses Kapitels) und besteht in folgendem: Die $x_1, \ldots, x_n$ werden als Ortskoordinaten von Teilchen angesehen, die sich im Potentialfeld U bewegen. Während der Bewegung sind Zusammenstöße mit anderen analogen Teilchen möglich. Das Stoßmodell kann weitgehend willkürlich sein, aber es gibt eine mittlere Weglänge, auf der kein Zusammenstoß erfolgt (freie Weglänge). Wenn sich im System ein statistisches Gleichgewicht eingestellt hat, dann hat die Verteilungsdichte der Koordinaten der

Teilchen im Phasenraum der Koordinaten und Geschwindigkeiten die Gestalt $C \exp(-E/\beta)$. Dabei sind C und β Konstanten, und die Energie E ist die Summe von potentieller und kinetischer Energie, also $E = U(x_1, \ldots, x_n) + T(\dot{x}_1, \ldots, \dot{x}_n)$ mit $\dot{x}_i = dx_i/dt$. Daraus ergibt sich wie früher die Dichte der gemeinsamen Verteilung der Ortskoordinaten $x_1, \ldots, x_n$ zu

$$p(x_1, \ldots, x_n) = \exp\left(-\frac{U}{\beta}\right) \Big/ \int dx_1 \cdots dx_n \exp\left(-\frac{U}{\beta}\right). \tag{6.4.5}$$

Aus (6.4.5) folgt, daß $p(x_1, \ldots, x_n)$ für solche x_i ein Maximum annimmt, die Lösungen des Systems (6.4.3) sind. Dabei kann der Anstieg in der Umgebung des Maximums durch die Wahl der Konstanten a_i beeinflußt werden. Das bedeutet, daß man beim Verfolgen der Trajektorien im Verlauf eines hinreichend großen Zeitintervalls Angaben über die Lage der Lösungen erhalten kann.

Ein einfaches Modell, das zu den gewünschten Ergebnissen führt, ist das folgende. Die Bewegungsgleichungen des Teilchens

$$\left.\begin{aligned} \frac{dp_i}{dt} &= \frac{\partial u}{\partial x_i} = -U'_{x_i}(x_1, \ldots, x_n), \\ \frac{dx_i}{dt} &= p_i, \qquad i = 1, \ldots, n, \end{aligned}\right\} \tag{6.4.6}$$

werden durch die entsprechenden Differenzengleichungen ersetzt. Dann wird ein beschränktes Gebiet im Raum der Koordinaten gewählt, beispielsweise ein n-dimensionaler Würfel mit hinreichend großen Seiten, der bequem in Teilgebiete zerlegbar ist. Die Anfangslage der Teilchen des Systems wird als gleichverteilt im gewählten Würfel angenommen. Weiter wird eine Größe $T_{\max}$ gewählt, die den maximal möglichen Wert der kinetischen Energie angibt. Alle Bewegungsrichtungen im Anfangsmoment seien gleichwahrscheinlich, und die Anfangsenergie betrage $T = \alpha T_{\max}$, wobei α wie immer eine auf $[0, 1]$ gleichverteilte Zufallsgröße bezeichnet. Die Geschwindigkeit im Anfangsmoment ergibt sich dann zu $\dot{x}_i = p_i = \sqrt{T} v_i$, wobei die v_i Komponenten eines in allen räumlichen Richtungen gleichverteilten Einheitsvektors sind. Die zeitliche Schrittweite hängt mit der freien Weglänge λ und der Wahrscheinlichkeit s eines Stoßes zusammen. Für Δt wählt man $\Delta t = s\lambda/\sqrt{T}$ (Weg durch Geschwindigkeit). Entsprechend den Gleichungen (6.4.6) werden die neuen Lagen der Teilchen und die neuen Geschwindigkeitskomponenten berechnet, wobei die Summe der Quadrate der Geschwindigkeitskomponenten die neue kinetische Energie des Systems liefert. Dann wird wieder die zeitliche Schrittweite bestimmt usw. Bei jedem Schritt wird geprüft, ob ein Zusammenstoß stattgefunden hat. Die Wahrscheinlichkeit eines Stoßes beträgt $s < 1$ für $T < T_{\max}$ und ist gleich Eins, wenn T nicht kleiner als $T_{\max}$ ist. Nach dem Zusammenstoß wird die Energie zu $\alpha_1 T_{\max}$ gewählt, die Bewegungsrichtung ist gleichverteilt wie zu Beginn der Teilchenbewegung. Daher übersteigt die kinetische Energie nicht $T_{\max}$, was zur Einstellung einer stationären Verteilung beiträgt.

Wenn die Teilchen das vorgegebene Teilgebiet verlassen, dann besteht Grund zur Annahme, daß sich dort keine Lösungen des Systems befinden. Solche Teilgebiete schließen wir aus der Betrachtung aus. Nun sei ein Teilgebiet festgestellt worden, in dem eine Konzentration der Teilchen stattfindet. Mit Hilfe der statio-

nären Verteilung wird dann

$$\mathsf{M}x_i = \frac{\int dx_1 \cdots \int dx_n\, x_i \exp(-U/\beta)}{\int dx_1 \cdots \int dx_n \exp(-U/\beta)}$$

berechnet, das Mittel über die Trajektorien (d. h. das Mittel der Werte in den Ecken des Polygonzuges) in den Grenzen dieses Teilgebietes. Dieser Mittelwert approximiert bei entsprechender Wahl der Konstanten a_i die i-te Komponente der Lösung, wenn es innerhalb des Teilgebietes eine Lösung gibt, oder die Summe der i-ten Komponenten der Lösungen, wenn es dort mehrere Lösungen gibt. Dabei streben für $T \to \infty$ alle $\mathsf{D}x_i$ gegen Null, wenn es nur eine Lösung gibt. Sind es mehrere, dann strebt wenigstens ein $\mathsf{D}x_i$ gegen eine von Null verschiedene Konstante. In diesem Fall muß man das Gebiet bezüglich der Veränderlichen teilen, für die $\mathsf{D}x_i$ maximal ist. Durch solche Unterteilungen kann man offensichtlich die Lösungen lokalisieren. Dabei sollte man beachten, daß die Werte von x_i in den Ecken des Polygonzuges abhängige Zufallsgrößen sind (Aufgabenstellung C in Kap. I). Die Verkleinerung von $T_{\max}$, λ und die Vergrößerung der a_j gibt uns die Möglichkeit, die Lösungen mit größerer Genauigkeit zu bestimmen.

Damit soll die Darstellung der grundlegenden Ideen in der von Tsudo und Kiyono [1] vorgeschlagenen Lösungsmethode für nichtlineare Gleichungssysteme abgeschlossen werden. Freilich wären noch weitere Untersuchungen nötig, um konkrete Vorschläge für eine günstige Wahl der Parameter $T_{\max}$, λ und a_i machen zu können. Es bleibt auch eine offene Frage, von welchem Moment an die Verteilung als stationär anzusehen ist oder wie man die Dispersionen der Größen x_i während der Rechnung effektiv schätzen kann. In der genannten Arbeit wird ein numerisches Beispiel angegeben. Es wird das System

$$xy - \frac{3}{16} = 0\,,$$

$$x^2 + y^2 - \frac{5}{8} = 0\,,$$

gelöst. Zunächst wurde $T_{\max} = 4$; $s = 0{,}1$; $\lambda = 0{,}1$; $a_i = 1$ gewählt. Nach 2000 Schritten wurde eine Trennung der Lösungen erreicht und $T_{\max}$ um den Faktor 410 verkleinert, $\lambda = 0{,}01$ und $a_i = 10$ gewählt. Durch weitere fünftausend Schritte wurden die Lösungen mit einem Fehler von etwa 5% bestimmt.

Die beschriebene Methode kann man offensichtlich auch bei der Extremwertbestimmung anwenden. Dabei kann man alle Extremwerte finden. Obwohl die Anwendung von Differentialgleichungen zur Lösung von Gleichungssystemen oder zur Extremwertbestimmung auch ohne Monte-Carlo-Methode möglich ist, schließt ein nichtstochastisches Verfahren die Schätzung der Dispersion und folglich eine Verteilung der Lösungen und Extremwerte aus. Dieser Umstand besitzt prinzipielle Bedeutung und sollte daher besonders betont werden.

3. Extremwertbestimmung einer Funktion. Die Aufgabe der Extremwertbestimmung wurde schon in Zusammenhang mit den Methoden der stochastischen Approximation und in Abschnitt 2 behandelt. Es gibt viele Arbeiten, in denen randomisierte Verfahren der Extremwertsuche betrachtet werden. Es genügt aber,

hier die diesen Fragen gewidmete Monographie von RASSTRIGIN [1] und die schon früher zitierte Arbeit von WAISBORD und JUDIN [1] zu nennen. Diese selbständige und vielversprechende Entwicklungsrichtung wird von uns nur teilweise im Zusammenhang mit anderen Aufgaben berührt. Hier werden wir nur zwei Bemerkungen zu Methoden der Extremwertsuche unter sehr allgemeinen Voraussetzungen bezüglich der Funktion $f(x)$ machen (Differenzierbarkeit der Funktion wird nicht vorausgesetzt).

Das einfachste Verfahren, das man sich denken kann, wenn die Funktion nicht als differenzierbar vorausgesetzt wird, besteht in folgendem (ZIELIŃSKI [1]): Gegeben sei eine Funktion $f(\boldsymbol{x})$ in einem beschränkten Gebiet $\mathcal{D}$ des euklidischen Raumes. Wir bilden eine Folge $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ von Realisierungen eines in $\mathcal{D}$ gleichverteilten Zufallsvektors und setzen dann

$$s_1 = f(\boldsymbol{x}_1)\,, \qquad s_i = \begin{cases} f(\boldsymbol{x}_i) & \text{für} \quad f(\boldsymbol{x}_i) > s_{i-1}\,, \\ s_{i-1} & \text{für} \quad f(\boldsymbol{x}_i) \leqq s_{i-1}\,. \end{cases} \tag{6.4.7}$$

$f(\boldsymbol{x})$ wird als eine im Lebesgueschen Sinn meßbare Funktion vorausgesetzt. Ist $s_0 = \sup f(\boldsymbol{x})$ (der Fall des Infimums wird analog betrachtet) und mes $\{\boldsymbol{x} : f(\boldsymbol{x}) = s_0\} = 0$, dann ist die Wahrscheinlichkeit dafür, daß ein x_i mit dem Extrempunkt x_0 zusammenfällt, gleich Null. Daher gibt uns das Verfahren (6.4.7) eigentlich keine Möglichkeit den Extrempunkt zu finden. Die Aufgabe bekommt einen Sinn, wenn ein Punkt aus der Menge $E^* = \{\boldsymbol{x} : f(\boldsymbol{x}) > s^*\}$ bestimmt werden soll, wobei E^* ein positives Maß hat. In diesem Fall kann das Verfahren (6.4.7) zu einem Erfolg führen. Zu den Aufgaben solcher Art gehört beispielsweise die Aufgabe über den optimalen Auslastungsplan einer Einrichtung (Kap. III, § 3). Leicht lassen sich weitere Beispiele für Aufgaben angeben, in denen ein hinreichend großer (oder kleiner) Wert von $f(\boldsymbol{x})$ gefunden werden soll, aber nicht unbedingt der Punkt, in dem $f(\boldsymbol{x})$ seinen größten (oder kleinsten) Wert erreicht. Ist jedoch $f(\boldsymbol{x})$ stetig, dann folgt bei hinreichend kleinem ε aus der Ungleichung mes $\{\boldsymbol{x} : f(\boldsymbol{x}) > s^*\} \leqq \varepsilon$, daß die Punkte der Menge E^* genügend nahe beim Supremumpunkt von $f(\boldsymbol{x})$ liegen. Wir bezeichnen mit $A(m = n; N)$ die Ereignisse, die darin bestehen, daß unter N Versuchen n Funktionswerte $f(\boldsymbol{x}_i)$ größer als der Wert s^* sind. Mit μ^* bezeichnen wir das Maß von E^*. Dann gilt

$$\mathsf{P}\big(A(m > 0; N) \mid \mu^* = \varepsilon\big) = 1 - \mathsf{P}\big(A(m = 0; N) \mid \mu^* = \varepsilon\big) = 1 - (1-\varepsilon)^N\,.$$

Wollen wir beispielsweise die Werte x aus der Menge E^* mit $\mu^* \leqq \varepsilon$ mit einer Wahrscheinlichkeit, die nicht kleiner als $1 - \varepsilon$ ist, erhalten, dann können wir die dafür notwendige Zahl N angeben, die der Ungleichung $(1 - \varepsilon)^N \leqq \varepsilon$ genügt. Daraus ergibt sich

$$N \log (1 - \varepsilon) \approx \log \varepsilon\,, \qquad N \approx \frac{\log \varepsilon}{\log (1 - \varepsilon)}\,.$$

In Tabelle 6.3 sind einige Werte für N in Abhängigkeit von ε angeführt.

ε	0,5	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
N	1	8	22	58	194	458	1057	3104	6903

Tabelle 6.3

Ein anderer Weg, der im Zusammenhang mit der Modellierung von Verteilungen steht, kann in gewissen Sonderfällen nützlich sein. Wir nehmen an, daß $|f(\boldsymbol{x})|$ für alle $\boldsymbol{x}$ aus $\mathcal{D}$ beschränkt ist. Man kann stets erreichen, daß $0 \leqq f(\boldsymbol{x}) \leqq M$ gilt und der Inhalt (das Maß) von $\mathcal{D}$ gleich Eins ist. Wenn $c_1 f(\boldsymbol{x})$ normiert ist und daher als Verteilungsdichte einer Zufallsgröße angesehen werden kann, dann werden sich die Stichprobenwerte für ξ um den Punkt gruppieren, in dem $f(\boldsymbol{x})$ seinen größten Wert hat. Ist auch noch die Dichte $c_n f^n(\boldsymbol{x})$ mit $n \gg 1$ einfach genug zu modellieren, so kann man hoffen, den größten Wert von $f(\boldsymbol{x})$ mit geringem Rechenaufwand zu finden. Man zeigt leicht, daß die v.-Neumannsche Methode in diesem Fall zu einem Verfahren führt, das dem eben beschriebenen (vgl. (6.4.7)) sehr ähnlich ist. Wenn jedoch für einige Veränderliche die Anwendung der Umkehrformeln oder anderer Methoden möglich ist, dann kann das der Ausgangspunkt für eine Verbesserung des Algorithmus (6.4.7) sein.

VII. MIT DER ZAHLENTHEORIE ZUSAMMENHÄNGENDE FRAGEN

§ 1. Gleichverteilte und vollständig gleichverteilte Folgen Der Begriff der arithmetischen Modellierung von Zufallsprozessen

Die stochastischen Methoden, die in den ersten sechs Kapiteln untersucht wurden, ermöglichen die Zurückführung von Aufgaben, die mit der Monte-Carlo-Methode lösbar sind, auf die Berechnung eines Integrals hoher Vielfachheit über einem Einheitswürfel. Das ergibt sich aus der von uns durchgeführten Konstruktion des Maßes, über das integriert wird. Der Integrand braucht dabei keine Glattheitseigenschaften zu besitzen und besitzt diese in der Regel auch nicht. Freilich zielen die Methoden der Dispersionsverkleinerung darauf hin, den Integrand möglichst zu einer Konstanten werden zu lassen. Wenn es gelingt, die Dispersion hinreichend klein zu machen, dann kann die Qualität der Zufallszahlen das Resultat nur wenig beeinflussen. Gerade im allgemeinen Fall erweist sich die Herleitung einer Pseudozufallsfolge als eng benachbart zum Problem der näherungsweisen Integration bei deterministischer Aufgabenstellung. Offensichtlich muß dabei die Funktionenklasse, für die die entsprechende Quadraturformel hergeleitet wird, hinreichend umfangreich sein. Prinzipiell ist es möglich, eine solche Punktfolge in einem Würfel beliebiger Dimension zu konstruieren, daß das arithmetische Mittel der Werte einer willkürlichen im Riemannschen Sinn integrierbaren Funktion in diesen Punkten bei Vergrößerung der Summandenzahl gegen den Wert des entsprechenden Integrals konvergiert. Dabei wird die Konvergenz im deterministischen Sinn verstanden. Dieses Resultat besitzt prinzipielle Bedeutung, und wir gehen anschließend näher darauf ein.

Wir betrachten nun eine unendliche Folge von Zahlen mit unendlich vielen Ziffern. In der Regel werden wir annehmen, daß die Zahlen als Dualzahlen aus dem Intervall $[0, 1)$ gegeben sind. Die Beschränkung auf ein Zahlensystem ist nicht wesentlich, vereinfacht aber manchmal die Darstellung.

Es seien ε_n, $n = 1, 2, \ldots$, Dualzahlen mit nichtbeschränkter Stellenzahl aus $[0, 1)$ und $\mathscr{E} = \varepsilon_1, \varepsilon_2, \ldots$ eine unendliche Folge dieser Zahlen. $\mathscr{E}$ heißt gleichverteilt im Weylschen Sinn auf $[0, 1)$, wenn für eine beliebige im Riemannschen Sinn eigentlich integrierbare Funktion h gilt:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} h(\varepsilon_n) = \int_0^1 h(x)\,dx. \tag{7.1.1}$$

Weiterhin führen wir die Bezeichnung

$$\mathsf{P}(S_n) = \lim_{N\to\infty} \frac{1}{N} \sum_{\substack{S_n \text{ erfüllt} \\ 1<n\leqq N}} 1$$

ein, wobei S_n eine Aussage in Abhängigkeit von einer natürlichen Zahl ist.

Aus (7.1.1) folgt, daß der Anteil der Glieder der Folge $\mathcal{E}$, die in das Intervall $[a, b)$, $0 \leqq a < b < 1$, fallen, für jedes a und b asymptotisch gleich der Länge $b - a$ dieses Intervalls ist, also

$$\mathsf{P}(a \leqq \varepsilon_n < b) = b - a \,. \tag{7.1.2}$$

Man kann auch zeigen, daß aus (7.1.2) wieder (7.1.1) folgt, so daß man (7.1.2) ebenso als Definitionsgleichung der Gleichverteilung nach WEYL benutzen kann.

Für die Konstruktion ist es bequemer, die Folge $\mathcal{E}$ als Funktion eines ganzzahligen Arguments aufzufassen, also $\varepsilon_n = f(n)$, $0 \leqq f(n) < 1$. Wenn dabei die Folge $\mathcal{E}$ gleichverteilt ist, so sagt man, daß auch $f(n)$ gleichverteilt ist. Für $f(n)$ wählt man häufig den gebrochenen Teil einer Funktion mit ganzzahligem Argument, also $f(n) = \{g(n)\}$. Die Funktion $g(n)$ kann beliebige reelle Werte annehmen. Ist $f(n)$ gleichverteilt und gilt $f(n) = \{g(n)\}$, so sagt man auch, daß $g(n)$ gleichverteilt ist nach dem Modul 1 (mod 1). Eine wichtige Bedeutung für das weitere hat das Weylsche Kriterium, das den Inhalt des anschließenden Satzes bildet.

Satz 7.1 (WEYL [1]). *Die Folge ε ist genau dann gleichverteilt, wenn*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \exp(2\pi i k \varepsilon_n) = 0$$

für jede von Null verschiedene ganze Zahl k gilt.

Beweis. Der Beweis der Notwendigkeit ist klar. Wir zeigen, daß der Satz auch hinreichend ist. Dazu gehen wir von (7.1.2) aus und nehmen ohne Beschränkung der Allgemeinheit an, daß ε_n der gebrochene Teil einer mod 1 gleichverteilten Funktion $g(n)$ ist.

Nun sei $G(t)$ eine Funktion mit der Periode 1, die auf dem Intervall $[0, 1)$ in folgender Weise definiert ist:

$$G(t) = \begin{cases} 1 & \text{für} \quad 0 \leqq t < \gamma \,, \\ 0 & \text{für} \quad \gamma \leqq t < 1 \,. \end{cases}$$

Dann gilt offenbar

$$\mathsf{P}\big(0 \leqq \{g(n)\} < \gamma\big) = \lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} G\big(g(n)\big) \,. \tag{7.1.3}$$

Wir wählen nun ein $\beta > 0$ so, daß gleichzeitig die Ungleichungen $2\beta < \gamma$ und $2\beta < 1 - \gamma$ erfüllt sind. Dann werden zwei stetige Funktionen G_1 und G_2 mit der Periode 1 folgendermaßen gebildet:

$$G_1(t) = \begin{cases} 1 & \text{für} \quad 0 \leqq t \leqq \gamma \,, \\ \text{linear} & \text{für} \quad \gamma \leqq t \leqq \gamma + \beta \,, \\ 0 & \text{für} \quad \gamma + \beta \leqq t \leqq 1 - \beta \,, \\ \text{linear} & \text{für} \quad 1 - \beta \leqq t \leqq 1 \,, \end{cases}$$

$$G_2(t) = \begin{cases} \text{linear} & \text{für} \quad 0 \leqq t \leqq \beta \,, \\ 1 & \text{für} \quad \beta \leqq t \leqq \gamma - \beta \,, \\ \text{linear} & \text{für} \quad \gamma - \beta \leqq t \leqq \gamma \,, \\ 0 & \text{für} \quad \gamma \leqq t \leqq 1 \,. \end{cases}$$

Offensichtlich gilt

$$G_2(t) \leqq G(t) < G_1(t) \,. \tag{7.1.4}$$

Sowohl $G_1(t)$ als auch $G_2(t)$ werden durch eine gleichmäßig konvergente Fourierreihe dargestellt:

$$G_1(t) = \gamma + \beta + \sum_{k=1}^{\infty} \big(a_k \exp(2\pi ikt) + b_k \exp(-2\pi ikt)\big) \,, \tag{7.1.5}$$

$$G_2(t) = \gamma - \beta + \sum_{k=1}^{\infty} \big(a'_k \exp(2\pi ikt) + b'_k \exp(-2\pi ikt)\big) \,. \tag{7.1.6}$$

Um den Beweis abzuschließen, muß man $t = g(n)$ setzen und die Reihen (7.1.5) und (7.1.6) für $n = 1, 2, \ldots, N$ addieren. Für $\beta \to 0$ folgt aus (7.1.3) und (7.1.4) das Geforderte.

Übrigens folgt aus dem Weylschen Kriterium sofort die Äquivalenz der Definitionen der Gleichverteilung nach (7.1.1) und (7.1.2).

Zur Herleitung eines Beispiels einer im Weylschen Sinn gleichverteilten Folge benötigt man das folgende einfache Lemma.

Lemma 7.1. *Es gilt die Ungleichung*

$$\left|\sum_{n=m+1}^{m+p} \exp(2\pi i\beta n)\right| \leqq \min\left(p, \frac{2}{|\sin(\pi\beta)|}\right). \tag{7.1.7}$$

Für ganzzahlige β ist das richtig. Ist β keine ganze Zahl, so kann man die Summenformel für die endliche geometrische Reihe anwenden, also

$$\sum_{n=m+1}^{m+\varrho} \exp(2\pi i\beta n) = \frac{\exp\big(2\pi i\beta(m+1)\big)\big(1 - \exp(2\pi i\beta p)\big)}{1 - \exp(2\pi i\beta)} \,.$$

Der Betrag des Zählers in diesem Ausdruck ist nicht größer als 2 und der Betrag des Nenners nicht kleiner als $|\sin(\pi\beta)|$, womit das Lemma bewiesen ist.

Folgerung. *Für jede ganze Zahl $k \neq 0$ und beliebige irrationale Zahl β gilt*

$$\lim_{N\to\infty} \frac{1}{N}\left|\sum_{n=1}^{N} \exp(2\pi ikn\beta)\right| \leqq \lim_{N\to\infty} \frac{1}{N} \min\left(N, \frac{1}{|\sin(\pi k\beta)|}\right) = 0 \,.$$

Daraus ergibt sich, daß die Folge

$$\varepsilon_n = \{n\,\beta\} \tag{7.1.8}$$

mit einer Irrationalzahl β im Weylschen Sinn gleichverteilt ist. Damit ist auch die Existenz gleichverteilter Folgen bewiesen.

Der Begriff der Gleichverteilung nach Weyl läßt sich auf Folgen von Vektoren erweitern. Man sagt, daß die Folge $Q = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ mit $\boldsymbol{x}_n = (\varepsilon_n^{(1)}, \ldots, \varepsilon_n^{(s)})$ im s-dimensionalen Einheitswürfel gleichverteilt ist, wenn für jede im Riemannschen Sinn eigentlich integrierbare Funktion $h(\boldsymbol{x})$

$$\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} h(x_n) = \int_0^1 dx^{(1)} \cdots \int_0^1 dx^{(s)}\, h(x) \tag{7.1.9}$$

gilt. Äquivalent dazu ist die Forderung

$$\mathsf{P}(a_i \leqq \varepsilon_n^{(i)} < b_i,\ i = 1, \ldots, s) = \prod_{i=1}^{s} (b_i - a_i)$$

für beliebige a_i und b_i mit $0 \leqq a_i < b_i \leqq 1$. Genau wie im eindimensionalen Fall betrachtet man oft Folgen von Vektoren, deren Komponenten Funktionen eines ganzzahligen Arguments sind, also $\boldsymbol{x}_n = \big(f_1(n), \ldots, f_s(n)\big)$, $0 \leqq f_i(n) < 1$. Es gibt auch ein mehrdimensionales Analogon zum Weylschen Kriterium.

Satz 7.2. *Genau dann ist die Folge Q im s-dimensionalen Einheitswürfel gleichverteilt, wenn die Beziehung*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \exp\big(2\pi i(k_1\varepsilon_n^{(1)} + \cdots + k_s\varepsilon_n^{(s)})\big) = 0$$

für alle $k_1, \ldots, k_s$ gilt, die nicht gleichzeitig Null sind.

Der Beweis des Satzes 7.2 entspricht völlig dem Beweis von Satz 7.1.

Die Komponenten der Vektoren der Folge Q können aus Elementen der Zahlenfolge $\mathscr{E}$ auf verschiedene Art gebildet werden. Beispielsweise kann man

$$\begin{aligned} &\boldsymbol{x}_1 = (\varepsilon_1, \ldots, \varepsilon_s)\,, \\ &\boldsymbol{x}_2 = (\varepsilon_{s+1}, \ldots, \varepsilon_{2s})\,, \\ &\ldots\ldots\ldots\ldots\ldots \\ &\boldsymbol{x}_n = (\varepsilon_{(n-1)s+1}, \ldots, \varepsilon_{ns})\,, \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{7.1.10}$$

oder

$$\begin{aligned} &\boldsymbol{x}_1 = (\varepsilon_1, \ldots, \varepsilon_s)\,, \\ &\boldsymbol{x}_2 = (\varepsilon_2, \ldots, \varepsilon_{s+1})\,, \\ &\ldots\ldots\ldots\ldots\ldots \\ &\boldsymbol{x}_n = (\varepsilon_n, \ldots, \varepsilon_{n+s+1})\,, \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{7.1.11}$$

setzen. Die Folge (7.1.11) nennen wir eine Kette der Folge $\mathscr{E}$ vom Rang s. Mit diesem Verfahren (7.1.11) zur Bildung einer Vektorfolge ist folgende Definition verbunden:

Definition. Die Folge $\mathscr{E}$ heißt *s-gleichverteilt,* wenn die durch eine Kette der Folge $\mathscr{E}$ vom Rang s gebildete Vektorfolge im s-dimensionalen Einheitswürfel gleichverteilt ist.

Wir notieren einige Eigenschaften einer s-gleichverteilten Folge.

1. Ist die Folge $\mathscr{E}$ s-gleichverteilt, so ist sie auch p-gleichverteilt für alle $p < s$. Die Aussage folgt unmittelbar aus dem mehrdimensionalen Weylschen Kriterium (Satz 7.2).

2. Die umgekehrte Aussage ist im allgemeinen falsch. Wir zeigen dazu, daß die Folge (7.1.8), $\beta, 2\beta, \ldots, n\beta, \ldots$, mit einer Irrationalzahl β nicht 2-gleichverteilt und folglich auch nicht p-gleichverteilt ist ($p > 2$). Für eine 2-gleichverteilte Folge

muß notwendigerweise die Beziehung

$$\mathsf{P}(\varepsilon_{n+1} < h(\varepsilon_n)) = \int_0^1 h(x)\,dx$$

für jede integrierbare Funktion h, $0 \leqq h < 1$, erfüllt sein. Speziell für $h(x) = x$ gilt

$$\mathsf{P}(\varepsilon_n < \varepsilon_{n+1}) = 1 - \mathsf{P}(\varepsilon_{n+1} < \varepsilon_n) = \frac{1}{2}.$$

Für die Folge (7.1.8) ist aber die Ungleichung $\varepsilon_n > \varepsilon_{n+1}$ gleichbedeutend mit $\varepsilon_n > \{\varepsilon_n + \beta\}$. Diese ist genau dann erfüllt, wenn $1 - \beta \leqq \varepsilon_n < 1$ gilt. Daraus ergibt sich

$$\mathsf{P}(\varepsilon_n > \varepsilon_{n+1}) = \mathsf{P}(1 - \beta < \varepsilon_n < 1) = \beta \neq \frac{1}{2}.$$

In diesem Fall war die Funktion $g(n)$, deren gebrochener Teil die Folge $\mathscr{E}$ bildet, eine lineare Funktion. WEYL [1] betrachtete auch kompliziertere Funktionen, insbesondere Polynome in n:

$$\varepsilon_n = \{\beta n^p + c_1 n^{p-1} + \cdots + c_p\}. \tag{7.1.12}$$

Dabei sind $c_1, \ldots, c_p$ beliebige Zahlen und β eine Irrationalzahl. Für diese Folgen gilt folgende Aussage, deren Beweis man in einer Arbeit von FRANKLIN [1] findet.

Satz 7.3. *Ist der Hauptkoeffizient β des Polynoms $g(n) = \beta n^p + c_1 n^{p-1} + \cdots + c_p$ eine Irrationalzahl, so ist die Folge* (7.1.12) *p-gleichverteilt, aber nicht $(p+1)$-gleichverteilt.*

Durch Wahl eines Polynoms $g(x)$ von hinreichend hohem Grad kann man also eine entsprechende hohe Ordnung der Gleichverteilung der Folge $\mathscr{E}$ sichern.

Der Grenzfall einer s-gleichverteilten Folge für beliebiges s wurde von N. M. KOROBOW [1] betrachtet.

Definition. Die Folge $\mathscr{E}$ heißt *vollständig gleichverteilt* oder im Korobowschen Sinn gleichverteilt, wenn sie für jedes natürliche s s-gleichverteilt ist.

Nach Satz 7.3 kann man erwarten, daß die Folge der gebrochenen Teile einer schneller als jede Potenz wachsenden Funktion eine vollständig gleichverteilte Folge bildet. Als erster konstruierte N. M. KOROBOW eine vollständig gleichverteilte Folge im Zusammenhang mit der Untersuchung der Verteilung der gebrochenen Teile einer Exponentialfunktion. Die Existenz vollständig gleichverteilter Folgen ist eine Aussage von prinzipieller Bedeutung, obwohl ihre algorithmische Erzeugung recht kompliziert ist.

Anschließend gehen wir kurz auf die Konstruktion dieser Folgen ein. Dabei schließt sich die Darstellung mit geringen Änderungen der Arbeit von POSTNIKOW [1] an[1]).

Es sei $\varepsilon = \delta_1, \delta_2, \ldots, \delta_N, \ldots$ eine Folge natürlicher Zahlen mit $0 \leqq \delta_i \leqq q - 1$ für alle i. Wir bilden nun eine Kette vom Rang s

$$(\delta_1, \delta_2, \ldots, \delta_s),\ (\delta_2, \delta_3, \ldots, \delta_{s+1}), \ldots, (\delta_N, \delta_{N+1}, \ldots, \delta_{N+s-1}), \tag{7.1.13}$$

[1]) Eine wesentlich einfachere Methode zur Konstruktion vollständig gleichverteilter Folgen, die aber nicht mit einer stochastischen Deutung verbunden ist, findet man in der Arbeit von KOROBOW [2].

deren Elemente wir Klammern nennen. Wir betrachten eine beliebige Klammer $\Delta^0 = (\delta_1^0, \delta_2^0, \ldots, \delta_s^0)$, die ihrerseits eine feste Menge von Zahlen zwischen 0 und $q - 1$ darstellt, und bezeichnen mit $P_N(\Delta^0)$ die Anzahl des Auftretens der Klammer Δ^0 bis zum N-ten Element der Folge.

Die Folge ε heißt eine normale Folge, wenn für jedes natürliche s und jedes feste Δ^0 die Beziehung

$$\lim_{N\to\infty} N^{-1} P_N(\Delta^0) = q^{-s}$$

erfüllt ist. Dann gilt folgender

Satz 7.4 (Schapiro-Pjatezki [1]). *Wenn eine Konstante C derart existiert, daß für beliebige natürliche s und jede feste Klammer Δ^0*

$$\lim_{N\to\infty} N^{-1} P_N(\Delta^0) < Cq^{-s}$$

gilt, dann ist die Folge ε normal.

Der Beweis des Satzes stützt sich auf folgendes Lemma.

Lemma 7.2. *Es sei r eine natürliche Zahl und $|\tau| \geqq 1$. Die Anzahl von Kombinationen aus l Elementen, in denen jedes feste Element $\delta^{(0)}$ in der Anzahl $l/q + \tau l/r$ vorkommt, übersteigt nicht $q^l r^4/(4l^2)$.*

Beweis. Die Anzahl von Kombinationen aus l Elementen, in denen $\delta^{(0)}$ genau k mal vorkommt, beträgt bekanntlich $\binom{l}{k}(q-1)^{l-k}$. Daraus ergibt sich die Anzahl der durch die Bedingungen des Lemmas bestimmten Kombinationen zu

$$\psi = \sum_{\left|k-\frac{l}{q}\right| \geq \frac{l}{r}} \binom{l}{k}(q-1)^{l-k} = q^l \sum_{\left|k-\frac{l}{q}\right| \geq \frac{l}{r}} \binom{l}{k}\left(\frac{1}{q}\right)^k\left(1-\frac{1}{q}\right)^{l-k}. \quad (7.1.14)$$

Das Weitere folgt aus bekannten Identitäten für die Binomialkoeffizienten (z. B. Natanson [1]):

$$\sum_{k=0}^{l} \binom{l}{k} x^k(1-x)^{l-k} = 1\,, \quad (7.1.15)$$

$$\sum_{k=0}^{l} (k-lx)^2 \binom{l}{k} x^k(1-x)^{l-k} = lx(1-x)\,. \quad (7.1.16)$$

Die erste Identität ist eine unmittelbare Folgerung der binomischen Formel. Die zweite ergibt sich folgendermaßen aus der ersten. Man differenziert die Gleichung $\sum_{k=0}^{l} \binom{l}{k} z^k = (1+z)^l$ nach z und multipliziert dann beide Seiten mit z. Von neuem wird nach z differenziert und wieder mit z multipliziert. Setzt man $z = x/(l-x)$, so erhält nacheinander

$$\sum_{k=0}^{l} \binom{l}{k} x^k(1-x)^{l-k} = 1\,, \quad (7.1.17)$$

$$\sum_{k=0}^{l} k \binom{l}{k} x^k(1-x)^{l-k} = lx\,, \quad (7.1.18)$$

$$\sum_{k=0}^{l} k^2 \binom{l}{k} x^k(1-x)^{l-k} = (lx + 1 - x)\, lx\,. \quad (7.1.19)$$

Nun multiplizieren wir (7.1.17), (7.1.18) und (7.1.19) jeweils mit l^2x^2, $-2lx$ und 1, addieren alles und erhalten (7.1.16).

Wir kommen nun zur Abschätzung der Größe ψ. Die Summation in (7.1.14) erstreckt sich über diejenigen k, die der Ungleichung

$$\left|\frac{k}{l} - \frac{1}{q}\right| \geqq \frac{1}{r} \qquad \text{oder} \qquad \left(k - \frac{1}{q}\right)^2 \left(\frac{r}{l}\right)^2 \geqq 1$$

genügen. Daher gilt

$$\sum_{\left|k - \frac{l}{q}\right| \geqq \frac{l}{r}} \binom{l}{k} \left(\frac{1}{q}\right)^k \left(1 - \frac{1}{q}\right)^{l-k}$$

$$\leqq \left(\frac{r}{l}\right)^2 \sum_{\left|k - \frac{l}{q}\right| \geqq \frac{l}{r}} \left(k - \frac{1}{q}\right)^2 \binom{l}{k} \left(\frac{1}{q}\right)^k \left(1 - \frac{1}{q}\right)^{l-k}.$$

Erstreckt man die Summation auf der rechten Seite der letzten Ungleichung über alle k, $0 \leqq k \leqq l$, so bleibt die Ungleichung richtig. Setzt man in (7.1.16) $x = \dfrac{1}{q}$, so ergibt sich das gewünschte Ergebnis, und das Lemma ist bewiesen.

Beweis des Satzes. Wir zerlegen die Folge in Gruppen von je l Elementen also $\delta_1, \ldots, \delta_l; \delta_{l+1}, \ldots, \delta_{2l}; \ldots$, und wählen eine feste Zahl r. Eine Gruppe soll zur Kategorie A gehören, wenn $\delta^{(0)}$ in der Anzahl $l/q + \tau l/r$, $|\tau| \leqq 1$, in dieser Gruppe vorkommt; andernfalls gehört die Gruppe zur Kategorie B. Mit $A(N)$ bzw. $B(N)$ bezeichnen wir die Anzahl der Gruppen in der Kategorie A bzw. B bei einer Folge der Länge $l[N/l]$. Dann kommt $\delta^{(0)}$ bis zum N-ten Element der Folge in folgender Anzahl vor:

$$\begin{aligned} P_N(\delta^{(0)}) &= A(N) \left(\frac{l}{q} + \frac{\tau l}{r}\right) + \vartheta_1 B(N)\, l + \vartheta_2 l \\ &= \left(\frac{N}{l} - B(N) + \vartheta_3\right)\left(\frac{l}{q} + \frac{\tau l}{r}\right) + \vartheta_1 B(N)\, l + \vartheta_2 l \\ &= \frac{N}{q} + \frac{N}{r}\tau + B(N)\left[\vartheta_1 l - \frac{l}{q} - \frac{\tau l}{r}\right] + \vartheta_3 \frac{l}{q} + \vartheta_3 \frac{\tau l}{r} + \vartheta_2 l \end{aligned}$$

mit $0 \leqq \vartheta_1, \vartheta_2 \leqq 1$ und $|\vartheta_3| \leqq 1$. Nach den Voraussetzungen des Satzes kommen für hinreichend große N Gruppen der Kategorie B nicht öfter als $2CN/q^l$-mal vor. Nach dem Lemma ist die Anzahl der verschiedenen Gruppen der Kategorie B von der Ordnung $O\left(q^l \dfrac{r^4}{l^2}\right)$. Daher gilt

$$B(N) < C_1 N q^l \frac{r^4}{l^2} \frac{1}{q^l} = C_1 N \frac{r^4}{l^2}.$$

Damit ergibt sich für $P_N(\delta^{(0)})$

$$\overline{\lim_{N \to \infty}} \left|\frac{1}{N} P_N(\delta^{(0)}) - \frac{1}{q}\right| \leqq \frac{1}{r} + O\left(\frac{r^4}{l^2}\right) + O\left(\frac{r^3}{l}\right).$$

Da r und l beliebig sind, erhalten wir $\lim P_N(\delta^{(0)})/N = 1/q$ für $N \to \infty$. Wenn man nun eine Klammer aus s Elementen als ein Element betrachtet (dann gibt es q^s Elemente), so ist auch leicht $\lim P_N(\Delta)/N = 1/q^s$ für $N \to \infty$ zu zeigen.

Wir betrachten nun ein von CHAMPERNOWNE [1] angegebenes Beispiel für eine normale Folge mit $q = 2$ Elementen. (Eine analoge Herleitung ist auch für beliebiges q möglich.)

Es sei $0, 1, \ldots, 2^r - 1$ eine Menge von Zahlen, die r Stellen in der Dualzahlendarstellung haben. Wir schreiben diese Zahlen nun in der natürlichen Reihenfolge hin, also

$$\text{für } r = 1\colon \quad \varrho_1 = 0, 1;$$

$$\text{für } r = 2\colon \quad \varrho_2 = 00, 01, 10, 11; \text{ usw.}$$

Daraus wird die Folge

$$\varrho_1 \varrho_2 \cdots \varrho_r \cdots \tag{7.1.20}$$

gebildet, in der die Kommas als Trennzeichen beibehalten werden. Wir zeigen nun, daß die Folge (7.1.20) normal ist. Dazu wird aus dieser Folge eine Kette vom Rang s gebildet und eine feste Klammer Δ_s gewählt. Wir setzen noch

$$S_r = \varrho_1 \varrho_2 \cdots \varrho_r \,. \tag{7.1.21}$$

Ist Δ_s so in ϱ_r enthalten, daß kein Komma Δ_s teilt, dann heißt Δ_s *ungeteilt*. Andernfalls heißt Δ_s *geteilt*. Für $s < r$ gibt es keine ungeteilten Klammern. Für $s \geqq r$ gibt es $s - r + 1$ Möglichkeiten, wie Δ_s in ϱ_r enthalten sein kann, weil das erste Element von Δ_s das erste Element einer r-stelligen Dualzahl oder auch das zweite Element usw. sein kann. Bei fester Lage von δ_s können die übrigen $s - r$ Elemente beliebig sein. Daraus folgt für die Anzahl der ungeteilten Klammern Δ_s in S_r der Ausdruck $(r - s + 1)\, 2^{s-r}$. Wird aber ein festes Komma betrachtet, dann kann es nicht mehr als s verschiedene Δ_s teilen. Daraus ergibt sich, daß δ_s in ϱ_r nicht öfter als $r \cdot 2^r$-mal geteilt vorkommen kann.

Wenn man mit h_r diejenige Anzahl bezeichnet, die angibt, wie oft Δ_s in ϱ_r vorkommt, so ergibt sich für $r \to \infty$

$$h_r = (r - s + 1)\, 2^{r-s} + O(2^r)\,.$$

Die Anzahl der Dualelemente in ϱ_r beträgt $r \cdot 2^r = y_r$, so daß

$$h_r = y_r 2^{-s} + O(y_r)$$

gilt. Für die Anzahl H_r des Auftretens von Δ_s in S_r ergibt sich

$$H_r = \sum_{i=1}^{r} h_i + O(r)\,.$$

Mit $x_r = \sum_{i=1}^{r} y_i$ erhält man $H_r = x_r 2^{-s} + O(x_r)$ oder für $H_r = H(x_r)$ als Funktion von x_r

$$\frac{H(x_r)}{x_r} = 2^{-s} + \frac{O(x_r)}{x_r}\,.$$

Mit $x_r \leqq x < x_{r+1}$ gilt

$$\frac{H(x)}{x} < \frac{H(x_{r+1})}{x_r} < \frac{H(x_{r+1})\, x_{r+1}}{x_{r+1} x_r}\,.$$

Wegen

$$\frac{x_{r+1}}{x_r} = \frac{x_r + (r + 1)\, 2^{r+1}}{x_r} \leqq \frac{1 + (r + 1)\, 2^{r+1}}{r 2^r} < 1 + 4$$

findet man schließlich

$$\overline{\lim} \frac{H(x)}{x} < 5 \cdot 2^{-s}.$$

Nach Satz 7.4 ist die Folge von CHAMPERNOWNE [1] eine normale Folge von Elementen.

Ferner sei ein System von Folgen

$$\begin{aligned} \mathscr{E}_1 &= \delta_{11}\delta_{12}\ldots, \\ \mathscr{E}_2 &= \delta_{21}\delta_{22}\ldots, \\ &\ldots\ldots\ldots\ldots \\ \mathscr{E}_k &= \delta_{k1}\delta_{k2}\ldots \end{aligned}$$

gegeben, wobei die Folge $\mathscr{E}_j$ aus den Elementen $0, 1, \ldots, q_{j-1}$ besteht, $j = 1, \ldots, k$; $q_j \geqq 2$. Für dieses System kann man eine Matrizenkette vom Rang s konstruieren:

$$\begin{pmatrix} \delta_{1,1} \ldots \delta_{1,s} \\ \ldots\ldots\ldots \\ \delta_{k,1} \ldots \delta_{k,s} \end{pmatrix}, \quad \begin{pmatrix} \delta_{1,2} \ldots \delta_{1,s+1} \\ \ldots\ldots\ldots \\ \delta_{k,2} \ldots \delta_{k,s+1} \end{pmatrix}, \ldots \tag{7.1.22}$$

Offensichtlich gibt es $(q_1q_2 \cdots q_k)^s$ verschiedene Matrizen, aus denen die Folge (7.1.22) bestehen kann.

Das System von Folgen (7.1.22) heißt gemeinsam normal, wenn die Anzahl $H_P(\Delta_s)$ des Auftretens einer beliebigen festen Matrix Δ_s in einer Folge (7.1.22) der Länge P für $P \to \infty$ gegen $(q_1q_2 \cdots q_k)^{-s}$ strebt.

Offenbar kann die Theorie der gemeinsam normalen Folgen auf die Theorie der normalen Folgen zurückgeführt werden, wenn man als Elemente der Folge die Spaltenvektoren

$$\begin{pmatrix} \delta_{1i} \\ \ldots \\ \delta_{ki} \end{pmatrix}, \qquad i = 1, 2, \ldots$$

nimmt und die Kette vom Rang s nach der oben genannten Regel bildet. Weil es $q_1q_2 \cdots q_k$ verschiedene Spalten gibt, ist in diesem Fall $q = q_1q_2 \cdots q_k$, und man kann einen dem Satz 7.4 entsprechenden Satz formulieren.

Satz 7.4'. *Wenn für das System von Folgen* (7.1.22) *eine Konstante C derart existiert, daß für jedes natürliche s und jede feste Matrix Δ_s*

$$\lim_{P\to\infty} \frac{H_P(\Delta_s)}{P} < C(q_1q_2 \cdots q_s)^{-s}$$

gilt, dann ist das System von Folgen gemeinsam normal.

Wir nehmen nun an, daß es gelungen ist, ein unendliches System von Folgen $\mathscr{E}_1, \mathscr{E}_2, \ldots$ so zu konstruieren, daß bei beliebigem festen natürlichem l das System $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_l$ gemeinsam normal ist. Die Elemente δ_{ij} jeder Folge $\mathscr{E}_i$ sollen die Werte $0, \ldots, q-1$ annehmen. Wir setzen nun

$$\alpha_j = \frac{\delta_{1j}}{q} + \frac{\delta_{2j}}{q^2} + \cdots = \sum_{i=1}^{\infty} \frac{\delta_{ij}}{q^i}.$$

Dann gilt

Satz 7.5. *Die Zahlenfolge $\alpha_1, \alpha_2, \ldots$ ist vollständig gleichverteilt.*

Beweis. Wir setzen $\boldsymbol{x}_k = (\alpha_k, \alpha_{k+1}, \ldots, \alpha_{k+s-1})$, $k = 1, 2, \ldots$, mit einer festen natürlichen Zahl s. Neben $\boldsymbol{x}_k$ betrachten wir den Punkt $\boldsymbol{x}_k^{(l)} = (\alpha_k^{(l)}, \ldots, \alpha_{k+s-1}^{(l)})$ mit

den durch die Teilsumme $\alpha_j^{(l)} = \sum_{i=1}^{l} \delta_{ij}/q^i$ dargestellten Koordinaten. Da wir das System $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_l$ als gemeinsam normal vorausgesetzt haben, kommt jeder Punkt $x_k^{(l)}$ mit der asymptotischen Häufigkeit q^{-ls} vor. Das bedeutet, daß bei Teilung des s-dimensionalen Einheitswürfels in q^{-ls} gleiche Würfelchen (jede Kante wird in q^l gleiche Teile zerlegt) die Punkte der Folge $x_1^{(l)}, x_2^{(l)}, \ldots$ eine solche asymptotische Häufigkeit besitzen, in eines der Würfelchen zu fallen, die gleich seinem Volumen ist. Da l und s beliebig waren, ist der Satz damit bewiesen.

Ein System von l gemeinsam normalen Folgen kann für jedes feste l leicht angegeben werden. Dazu kann man beispielsweise die Folge von CHAMPERNOWNE für $q = 2^l$ konstruieren und jedem Element einen Spaltvektor aus l Nullen und Einsen zuordnen. Ein unendliches System ist nicht so leicht angebbar. Dafür wäre es günstig, ein Induktionsverfahren zu finden, so daß man durch „Hinzufügen" einer geeigneten Folge zu einem normalen System wieder ein gemeinsam normales System erhält. Anschließend werden wir ein solches sukzessives Verfahren beschreiben (STARTSCHENKO [1], [2]).

Es sei $\delta_1\delta_2 \ldots$ die ursprüngliche Folge und Δ_s eine beliebige feste Klammer einer Kette vom Rang s von dieser Folge. Mit $H_r(\Delta_s)$ bezeichnen wir die Anzahl der Klammern Δ_s auf einem Abschnitt aus den r Klammern, die die Kette bilden. Wegen der Normalität der ursprünglichen Folge gibt es ein $R = R(s, \delta_s)$, so daß für $r' > R$ und $r'' > R$

$$\left|\frac{H_{r'}(\Delta_s)}{r'} - \frac{H_{r''}(\Delta_s)}{r''}\right| < q^{-s} \tag{7.1.23}$$

gilt. Wegen

$$\frac{H_{r'}(\Delta_s) - H_{r''}(\Delta_s)}{r' - r''} = \frac{\dfrac{H_{r'}(\Delta_s)}{r'} - \dfrac{H_{r''}(\Delta_s)}{r''}}{1 - \dfrac{r''}{r'}} + \frac{\dfrac{H_{r''}(\Delta_s)}{r''} - \dfrac{H_{r''}(\Delta_s)}{r'}}{1 - \dfrac{r''}{r'}} \tag{7.1.24}$$

folgt für $r'' > \tilde{R}(\Delta_s)$ mit geeignetem $\tilde{R}(\Delta_s)$ aus der Normalität der Ausgangsfolge die Ungleichung

$$\frac{H_{r''}(\Delta_s)}{r''} < C_1 q^{-s}\,.$$

Wir wählen nun $N(s) = \sup\,(\tilde{R}(\Delta_s), R(s, \Delta_s))$. Dann ergibt sich bei beliebigem s für $r'' > N(s)$ und $r' - r'' > r''$ aus (7.1.23) und (7.1.24) die Beziehung

$$\begin{aligned}\frac{H_{r'}(\Delta_s) - H_{r''}(\Delta_s)}{r' - r''} &< \frac{r'q^{-s}}{r' - r''} + H_{r''}(\Delta_s)\,\frac{\dfrac{1}{r''} - \dfrac{1}{r'}}{1 - \dfrac{r''}{r'}} \\ &= q^{-s}\left(1 + \frac{r''}{r' - r''}\right) + \frac{H_{r''}(\Delta_s)}{r''} \\ &< q^{-s}\left(1 + \frac{r''}{r' - r''} + C_1\right) < Cq^{-s}\,.\end{aligned}$$

Ferner setzen wir

$$\overline{N}(2) = N(2) \qquad \text{und} \qquad \overline{N}(2s) = 2^{n_s}\overline{N}(2s - 2) + 2^{n_s} - 1\,,$$

wobei n_s eine beliebige natürliche Zahl mit

$$N(2s) \leqq 2^{n_s}\overline{N}(2s - 2) + 2^{n_s} - 2$$

ist. Dann gilt also $\overline{N}(2s) > N(2s)$ für $s > 1$.

Eine mit der Ausgangsfolge gemeinsam normale Folge konstruieren wir nun folgendermaßen in Gruppen, indem wir der Ausgangsfolge jeweils darunterstehend eine zweite Reihe zuordnen.

$$\text{0. Gruppe} \quad \begin{cases} \delta_1 \cdots \delta_{N(2)-1}\,, \\ \delta_1 \cdots \delta_{N(2)-1}\,, \end{cases}$$

$$\text{1. Gruppe} \quad \begin{cases} \delta_{N(2)} \quad \cdots \delta_{2N(2)}\,, \quad \delta_{2N(2)+1} \cdots\,, \ \delta_{4N(2)+2} \cdots \delta_{\overline{N}(4)-1}\,, \\ \delta_{N(2)+1} \cdots \delta_{2N(2)+1}, \ \delta_{2N(2)+2} \cdots\,, \ \delta_{4N(2)+3} \cdots \delta_{\overline{N}(4)}\,, \end{cases}$$

$$\text{2. Gruppe} \quad \begin{cases} \delta_{\overline{N}(4)} \quad \cdots \delta_{2\overline{N}(4)}\,, \quad \delta_{2\overline{N}(4)+1} \cdots \delta_{4\overline{N}(4)+2}\,, \ \delta_{4\overline{N}(4)+3} \cdots \delta_{\overline{N}(6)-1}\,, \\ \delta_{\overline{N}(4)+2} \cdots \delta_{2\overline{N}(4)+2}\,, \ \delta_{2\overline{N}(4)+3} \cdots \delta_{4\overline{N}(4)+4}\,, \ \delta_{4\overline{N}(4)+5} \cdots \delta_{\overline{N}(6)+1}\,. \end{cases}$$

Wir beweisen nun, daß wir so zwei gemeinsam normale Folgen erhalten haben. Dazu stellen wir zwei feste Klammern Δ_k und Δ_k' in einer Matrix

$$D_k = \begin{pmatrix} \Delta_k \\ \Delta_k' \end{pmatrix}$$

zusammen. Wir setzen nun $x = 2^n\overline{N}(2s) + 2^n - 2$, $n = 0, \ldots, n_{s+1} - 1$, und bezeichnen mit $G(x)$ die Anzahl des Auftretens von D_k in einer Kette vom Rang k und der Länge x im System (7.1.21). Dann wählen wir $s \geqq k$ und bezeichnen mit L die Anzahl des Auftretens von D_k bis zum Glied mit dem Index $\overline{N}(2k) - 1$. Das Auftreten von Δ_k in einer Gruppe mit dem Index s bedeutet, daß an den entsprechenden Stellen der Kette vom Rang $2s$ von der Ausgangsfolge eine der $q^{2(s-k)}$ Kombinationen der Art

$$\underbrace{\Delta_k \cdots}_{s \text{ Elemente}} \quad \underbrace{\Delta_k' \cdots}_{s \text{ Elemente}}$$

erscheint (möglicherweise außer den $2s$ letzten Gliedern der Gruppe). Auf jedem durch ein Komma geteilten Abschnitt tritt diese Kombination weniger als CRq^{-2s}-mal auf, wobei R die Anzahl der Elemente in diesem Abschnitt bedeutet. Die Anzahl solcher Kombinationen beträgt $q^{2(s-k)}$, daher kommt Δ_k nicht öfter als $q^{2(s-k)}CRq^{-2s} = CRq^{-2k}$-mal vor. Dann gilt $G(x) \leqq L + Cxq^{-2k} + O(s^2)$, wo bei der Summand durch den Übergang von einer Gruppe zur anderen entsteht. Damit ergibt sich

$$\frac{G(x)}{x} \leqq \frac{L}{x} + Cq^{-2k} + O\left(\frac{s^2}{x}\right)$$

mit $s^2/x \to 0$, da $x > C_2 2^s$ mit einer Konstanten C_2 gilt. Also gilt $\lim G(x)/x < Cq^{-2k}$ für $x \to \infty$, und unsere Behauptung ist bewiesen, wenn die Bedingungen des Satzes 7.4 erfüllt sind. Das wird aber genauso geprüft, wie das beim Beweis der Normalität der Folge von Champernowne geschah.

Nimmt man also beispielsweise als Ausgangsfolge die Folge von Champernowne, so kann man ein unendliches System gemeinsam normaler Folgen konstruieren und folglich eine vollständig gleichverteilte Zahlenfolge.

Eine vollständig gleichverteilte Folge bildet eine universelle Folge, mittels derer man Riemannsche Integrale beliebiger Vielfachheit (darunter abzählbar unendlicher) berechnen kann. Es ist zu bemerken, daß es für beliebige im Lebesgueschen Sinn integrierbare Funktionen eine solche universelle Folge nicht gibt. Das folgt aus der einfachen Tatsache, daß eine durch eine diskrete Punktfolge gebildete

Menge das Lebesguesche Maß Null hat. Daher gibt es immer eine im Lebesgueschen Sinn integrierbare Funktion, die auf dieser Punktmenge gar nicht definiert ist.

Offensichtlich muß die Zahlenfolge, die wir durch die Realisierungen einer auf $[0, 1)$ gleichverteilten Zufallsgröße bilden und für die Berechnung einer hinreichend umfangreichen Aufgabenklasse benutzen wollen, Eigenschaften besitzen, die im gewissen Sinn den Eigenschaften einer vollständig gleichverteilten Folge nahekommen. Die Natürlichkeit dieser Forderung wird auch durch folgende Überlegungen unterstrichen: Aus dem starken Gesetz der großen Zahlen folgt, daß die Folge der Realisierungen einer auf $[0, 1)$ gleichverteilten Zufallsgröße mit der Wahrscheinlichkeit 1 eine vollständig gleichverteilte Folge ist. Es sei nämlich $\alpha = \alpha_1, \alpha_2, \ldots$ eine solche Folge und $(\alpha_1, \ldots, \alpha_s), (\alpha_2, \ldots, \alpha_{s+1}), \ldots$ eine Kette vom Rang s. Dann gilt für jede im Riemannschen Sinn über dem s-dimensionalen Einheitswürfel $\mathcal{D}_s$ integrierbare Funktion mit der Wahrscheinlichkeit 1

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\alpha_i, \ldots, \alpha_{s+1}) = \int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{7.1.25}$$

da die Klammern der Kette Realisierungen von nicht mehr als s abhängigen Zufallsgrößen darstellen (Kap. I, § 3). Schließlich ist im Zusammenhang mit dem Begriff der vollständigen Gleichverteilung folgende von N. N. Tschenzow [2] betrachtete Aufgabe von unmittelbarem Interesse.

Gegeben ist eine homogene Markoffsche Kette mit einer endlichen Anzahl von Zuständen γ_i, $i = 1, \ldots, n$, und der Übergangsmatrix $\mathcal{P} = ((p_{ij}))$. Mit dem schon mehrfach beschriebenen Verfahren wird nun die Markoffsche Kette durch eine Zahlenfolge $\mathcal{E} = \varepsilon_1, \varepsilon_2, \ldots$ $(0 \leqq \varepsilon_i < 1)$ modelliert, d. h., man modelliert die Folge der diskreten Verteilungen nach Formel (2.2.6), wobei man als Zufallszahlen die Zahlen der Folge $\mathcal{E}$ wählt. Weiterhin werden wir dieses Modellierungsverfahren als Standardverfahren bezeichnen. Wir setzen noch voraus, daß die Matrix $\mathcal{P}$ ergodisch ist, es existiert also eine eindeutige stationäre Wahrscheinlichkeitsverteilung $\{q_i\}$, $i = 1, \ldots, n$. Wie schon früher bemerkt wurde, kann man sich bei der Berechnung des Mittelwerts $\sum_{i=1}^{n} q_i f(\omega_i)$ über die unbekannte stationäre Verteilung der Kette auf die Modellierung einer einzigen Trajektorie beschränken, die nur von hinreichend großer Länge sein muß. Unsere Aufgabe besteht nun in folgendem: Welche Beschränkungen ergeben sich für die Folge $\mathcal{E}$ aus der Forderung, daß für eine beliebige Markoffsche Kette mit ergodischer Matrix und beliebiger Anfangsverteilung bei der Modellierung mit dem Standardverfahren für jede Wahl der Funktion $f(\omega)$ die Beziehung

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(\gamma_{i_k}) = \sum_{i=1}^{N} q_i f(\omega_i) \tag{7.1.26}$$

erfüllt ist? Dabei bilden die γ_{i_k}, $k = 1, \ldots, N$, die Folge von Zuständen der Kette, die sich durch die Modellierung ergeben. Die Lösung dieser Aufgabe liefert der folgende Satz, der hier ohne Beweis angegeben wird.

Satz 7.6. *Genau dann erhält man mit dem Standardverfahren bei der Modellierung einer beliebigen Markoffschen Kette mit endlich vielen Zuständen und eindeutiger*

stationärer Wahrscheinlichkeitsverteilung eine ergodische Trajektorie im Sinn von (7.1.26) (*schwach ergodisch*), *wenn die Folge $\mathcal{E}$ vollständig gleichverteilt ist.*

Die Konstruktion normaler, gemeinsam normaler und vollständig gleichverteilter Folgen in dem oben beschriebenen Sinn bildet ein arithmetisches Modell einer gleichverteilten Zufallsgröße. Wie wir sahen, ist ein Modell solcher Art recht kompliziert, und gewöhnlich muß man sich auf gröbere und angenäherte Modelle beschränken (mehr dazu in § 2).

Demgegenüber ist es günstig, daß sich die Modellierung einer verhältnismäßig komplizierten Erscheinung, wie sie die Brownsche Bewegung darstellt, recht einfach verwirklichen läßt. Ergebnisse zur arithmetischen Modellierung der Brownschen Bewegung wurden von I. P. Kubiljus und J. W. Linnik [1] erhalten. Dieses Modell wollen wir nun kurz beschreiben.

Es sei p eine Primzahl, $p > 2$. Das Legendresche Symbol (m/p) der Zahl m in bezug auf p ist gleich 1, wenn m quadratischer Rest ist, gleich 0 für $m = 0$ und gleich -1, wenn m quadratischer Nichtrest ist. Das Symbol (m/p) ist also bei festem p eine periodische Funktion mit der Periode p des ganzzahligen Arguments m, definiert durch

$$\left(\frac{m}{p}\right) = \begin{cases} 0 & \text{für} \quad m = 0\,, \\ 1 & \text{für} \quad m = 1, 2^2, \ldots, \left(\dfrac{p-1}{2}\right)^2, \\ -1 & \text{sonst}\,. \end{cases}$$

Ferner sei R eine ungerade quadratfreie Zahl und $R = p_1 p_2 \cdots p_s$ die zugehörige Primfaktorenzerlegung.

Das Jacobische Symbol wird dann durch

$$\left(\frac{m}{R}\right) = \left(\frac{m}{p_1 \cdots p_s}\right) = \left(\frac{m}{p_1}\right) \cdots \left(\frac{m}{p_s}\right) = \prod_{R/p} \left(\frac{m}{p}\right)$$

definiert, wobei die Faktoren auf der rechten Seite Legendresche Symbole sind (*R/p bedeutet, daß p alle Primteiler von R durchläuft*). Das Legendresymbol ist daher ein Sonderfall des Jacobischen Symbols. Das gesuchte Modell wird nun mit Hilfe der Summen

$$S_R(m, s, t; h) = \frac{1}{\sqrt{h}} \sum_{hs < n \leqq ht} \left(\frac{m+n}{R}\right), \qquad 0 \leqq s < t\,,$$

gebildet. Dann gilt folgender Satz:

Satz 7.7. *R soll eine beliebige wachsende unendliche Folge ungerader quadratfreier Zahlen durchlaufen, so daß für jedes feste $c \geqq 0$*

$$\prod_{R/p} \left(1 - \frac{c}{p}\right) \to 1$$

und $h = h(R) \to \infty$, $\log h / \log R \to 0$ für $R \to \infty$ gilt. Dann erhält man

$$\mathsf{P}\big(S_R(m, s, t; h) < x\big) \to \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{x} e^{-u^2/2(t-s)}\, du$$

und bei beliebiger Wahl disjunkter Intervalle

$$(s_1, t_1,), \ldots, (s_k, t_k)\,, \qquad 0 < s_j < t_j\,, \qquad j = 1, \ldots, k\,,$$

$$\mathsf{P}(S_R(m, s_1, t_1; h) < x_1, \ldots, S_R(m, s_k, t_k; h) < x_k)$$

$$\to \prod_{j=1}^{k} \lim_{R\to\infty} \mathsf{P}(S_R(m, s_j, t_j; h) < x_j)\,.$$

Dieser Satz beschreibt eigentlich eine Familie von Modellen der Brownschen Bewegung. Kubiljus und Linnik haben solche Modelle aufgestellt, die mit den Restsymbolen höherer Potenzen verknüpft sind. Die Frage danach, welche der Modelle in numerischer Beziehung am günstigsten sind, bedarf noch einer genaueren Untersuchung.

§ 2. Über lineare Rekursionsformeln zur Erzeugung von Pseudozufallszahlen

1. Folgen von Zahlen mit beliebig vielen Ziffern. In diesem Abschnitt werden wir uns mit Folgen beschäftigen, die durch eine eindeutige rekursive Beziehung erzeugt werden. Zunächst betrachten wir eine Folge $\mathscr{E} = \varepsilon_1, \varepsilon_2, \ldots$ von Zahlen mit beliebig vielen Ziffern und nehmen an, daß diese Folge rekursiv durch die Formel

$$\varepsilon_{n+m} = F(\varepsilon_{n+m-1}, \ldots, \varepsilon_n)\,, \qquad n = 0, 1, \ldots\,, \tag{7.2.1}$$

mit festem $m \geqq 1$ erzeugt wird. Natürlich wird man versuchen ein solches F zu wählen, daß die Folge $\mathscr{E}$ in gewissem Sinn die Eigenschaften der Gleichverteilung erhält. Aber schon das vollständige Hinschreiben der Bedingungen, die durch die Forderung der Gleichverteilung an F zu stellen sind, ist offensichtlich recht schwierig, und wir werden eine solche Fragestellung nicht weiter verfolgen. Wir geben nur noch einen einfachen Satz an (Franklin [1]).

Satz 7.8 *Es sei* $\{\varepsilon_n\}$ *eine Folge auf* $[0, 1)$ *mit*

$$\varepsilon_{n+1} = F(\varepsilon_n)\,, \qquad n = 0, 1, \ldots$$

Diese Folge kann nicht p-gleichverteilt sein für beliebiges $p > 1$, *wenn* $F(x)$ *wenigstens einen Stetigkeitspunkt auf* $[0, 1)$ *besitzt.*

Beweis. Es sei $F(x)$ in $x = a$ stetig, $a \in [0, 1)$ und $F(a) = b$. Dann gibt es ein $\delta > 0$ so, daß $|F(x) - b| < 1/4$ für $|x - a| < \delta$ gilt. Wir bezeichnen nun mit I den Durchschnitt der Menge $\{x : |x - a| < \delta\}$ mit dem Intervall $[0, 1)$ und mit $\tilde{I}$ ein beliebiges Intervall auf $[0, 1)$, dessen Punkte von b um mehr als $1/4$ entfernt sind. Dann ist leicht zu sehen, daß niemals ein Punktepaar $\varepsilon_n, \varepsilon_{n+1}$ der Folge in dem Intervall $I \times \tilde{I}$ liegen kann. Damit ist der Satz bewiesen.

Wir werden nun den einfachsten Fall einer Funktion $F(x)$ untersuchen, wenn $F(x)$ die Gestalt

$$F(x) \equiv \{Mx + \vartheta\} \tag{7.2.2}$$

mit einer natürlichen Zahl $M > 1$ bei beliebigem reellem ϑ hat. Die Funktion $\{Mx + \vartheta\}$ ist in allen Punkten von $[0, 1)$ mit Ausnahme einer endlichen Anzahl von M oder $M - 1$ Punkten stetig. Es ist auch leicht zu zeigen, daß für die Folge $\varepsilon_{n+1} = \{M\varepsilon_n\}$ die Beziehung $\mathsf{P}(\varepsilon_n > \varepsilon_{n+1} > \varepsilon_{n+2}) = (1 - M^{-1})/6$ gilt. Für eine 2-Gleichverteilung müßte diese Wahrscheinlichkeit aber $1/6$ betragen. Daher bildet eine rekursive Folge, die durch den gebrochenen Teil einer linearen Funktion definiert ist, nicht einmal eine 2-gleichverteilte Folge. Dennoch müssen alle Versuche, diesen Sachverhalt durch Wahl einer komplizierteren Funktion F zu ändern, zu einem Mißerfolg führen. Im Zusammenhang damit geben wir ohne Beweis folgenden Satz an (FRANKLIN [1]):

Satz 7.9. *Es sei $p(x)$ ein Polynom mit reellen Koeffizienten. Wir setzen voraus, daß für ein gewisses ε_0 die nach der Formel $\varepsilon_{n+1} = \{p(\varepsilon_n)\}$ konstruierte Folge der ε_n auf $[0, 1)$ gleichverteilt ist. Dann ist entweder $p(x) = x + \beta$ mit einer Irrationalzahl β oder $p(x) = Mx + \vartheta$ mit einer natürlichen Zahl $M > 1$.*

Die Zahlen ε_0, für die bei jedem natürlichen $M > 1$ die durch $\varepsilon_{n+1} = \{M\varepsilon_n\}$ definierte Folge der ε_n auf $(0, 1)$ gleichverteilt ist, heißen absolut normal. Die Menge der auf $(0, 1)$ absolut normalen Zahlen hat das Lebesguesche Maß Eins. Zuerst wurden solche Zahlen im Jahr 1917 von LEBESGUE und SIERPIŃSKI konstruiert. Es ist klar, daß die durch

$$\varepsilon_{n+1} = \{M\varepsilon_n + \vartheta\}\,, \qquad \vartheta > 0\,, \tag{7.2.3}$$

definierte Folge analoge Eigenschaften besitzt.

Weiterhin wird es sich zeigen, daß die Folge (7.2.3) für hinreichend großes M viele Eigenschaften hat, wie sie einer Folge von Realisierungen einer auf $(0, 1)$ gleichverteilten Zufallsgröße zukommen. Um das zu zeigen, benötigen wir das folgende von SOBOL [8] gefundene Resultat.

Es sei $f(\boldsymbol{x})$ eine im s-dimensionalen Einheitswürfel $\mathcal{D}_s$ definierte und dort im Riemannschen Sinn integrierbare Funktion. Wir setzen

$$\varphi_M(x) = f(x, \{M_1 x\}, \{M_1 M_2 x\}, \ldots, \{M_1 \cdots M_{s-1} x\})\,,$$

wobei $M_1, M_2, \ldots, M_{s-1}$ natürliche Zahlen sind und $M = \min(M_1, \ldots, M_{s-1}) < 1$ gilt. Mit den Bezeichnungen

$$J = \int_{\mathcal{D}_s} f(\boldsymbol{x})\, d\boldsymbol{x} \qquad \text{und} \qquad J_M = \int_0^1 \varphi_M(x)\, dx$$

gilt der folgende Satz.

Satz 7.10. *Sind unter den angegebenen Voraussetzungen für ein gewisses $M_0 \geqq 2$ und alle $M > M_0$ alle $\varphi_M(x)$ im Riemannschen Sinn über $(0, 1)$ integrierbar, dann gilt*

$$\lim_{M \to \infty} J_M = J\,.$$

Beweis. Wir zerlegen das Intervall $(0, 1)$ in M_1 gleiche Intervalle und entsprechend J_M in eine Summe von M_1 Integralen. In jedem Teilintegral wird die

Substitution $M_1 x = j_1 + y_1$ durchgeführt. Damit erhält man

$$J_M = \sum_{j_1=0}^{M_1-1} \int_{j_1/M_1}^{(j_1+1)/M_1} \varphi_M(x)\,dx$$

$$= \frac{1}{M_1} \sum_{j_1=0}^{M_1-1} \int_0^1 f\left(\frac{j_1+y_1}{M_1}, y_1, \ldots, \{M_2 \cdots M_{s-1} y_1\}\right) dy_1 .$$

In gleicher Weise wird jedes der sich ergebenden Integrale in eine Summe von M_2 Integralen zerlegt und in jedem Teilintegral die Substitution $M_2 y_1 = j_2 + y_2$ durchgeführt. Dann erhält man

$$J_M = \frac{1}{M_1 M_2} \sum_{j_1=0}^{M_1-1} \sum_{j_2=0}^{M_2-1} \int_0^1 f\left(\frac{j_1+y_1}{M_1}, \frac{j_2+y_2}{M_2}, y_2, \ldots, \{M_3 \cdots M_{s-1} y_2\}\right) dy_2 .$$

Wenn man diesen Prozeß s-mal wiederholt (beim letzten Mal wird jedes Integral in M Teile zerlegt), so ergibt sich

$$J_M = \frac{1}{M_1 \cdots M_{s-1} M} \sum_{j_1=0}^{M_1-1} \cdots \sum_{j_{s-1}=0}^{M_{s-1}-1} \sum_{j_s=0}^{M-1} \int_0^1 f\left(\frac{j_1+y_1}{M_1}, \ldots, \frac{j_{s-1}+y_{s-1}}{M_{s-1}}, \frac{j_s+y_s}{M}\right) dy_s .$$

Nun seien $\overline{m}(j_1, \ldots, j_s)$ und $\underline{m}(j_1, \ldots, j_s)$ die obere und die untere Grenze der Funktion f im Parallelepiped

$$\frac{j_k}{M_k} \leqq x_k \leqq \frac{j_k+1}{M_k}, \qquad k = 1, \ldots, s, , \qquad M_s = M .$$

Offensichtlich gilt

$$\sum_{j_1=0}^{M_1-1} \cdots \sum_{j_s=0}^{M-1} \underline{m}(j_1, \ldots, j_s)\,(M_1 \cdots M_{s-1} M)^{-1} \leqq J_M$$

$$\leqq \sum_{j_1=0}^{M_1-1} \cdots \sum_{j_s=0}^{M-1} \overline{m}(j_1, \ldots, j_s)\,(M_1 \cdots M_{s-1} M)^{-1} .$$

In dieser Ungleichung steht rechts und links die obere bzw. untere Darbouxsche Summe für f in $\mathcal{D}_s$. Daher strebt J_M für $M \to \infty$ gegen J.

Franklin [1] zeigte 1963, daß die Folge (7.2.3) für $M \to \infty$ vollständig gleichverteilt ist. Wir werden hier den Beweis einer etwas allgemeineren von Sobol [8] gefundenen Aussage angeben. Dazu führen wir die Bezeichnung

$$\boldsymbol{x}_k = (\varepsilon_k, \ldots, \varepsilon_{k+s-1}) \qquad \text{und} \qquad \tilde{\boldsymbol{x}}_k = (\varepsilon_{ks+1}, \ldots, \varepsilon_{(k+1)s})$$

ein und nehmen $M_1 = M_2 = \cdots = M_{s-1} = M$ an. Dann gilt folgender Satz.

Satz 7.11. *Ist $f(\boldsymbol{x})$ im Riemannschen Sinn über $\mathcal{D}_s$ integrierbar und sind für ein gewisses $M_0 \geqq 2$ und alle $M > M_0$ alle φ_M im Riemannschen Sinne über* (0, 1)

integrierbar, dann gilt für fast alle ε_0 aus (0, 1)

$$\lim_{M\to\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} f(\boldsymbol{x}_k) = J \qquad \textit{und} \qquad \lim_{M\to\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} f(\tilde{\boldsymbol{x}}_k) = J\,.$$

Beweis. Es sei ε_0 eine absolut normale Zahl. Dann gilt

$$f(\boldsymbol{x}_k) = f(\{M^{k+1}\varepsilon_0\}, \ldots, \{M^{k+s}\varepsilon_0\}) = \varphi_M(\{M^{k+1}\varepsilon_0\}) = \varphi_M(\varepsilon_{k+1})\,.$$

Da die Folge $\mathscr{E}$ gleichverteilt ist, ergibt sich daraus die erste Aussage des Satzes, d. h. das erwähnte Resultat von Franklin.

Nun setzen wir $M^s = \tilde{M}$. Dann gilt

$$f(\tilde{\boldsymbol{x}}_k) = f(\{M\tilde{M}^k\varepsilon_0\}, \ldots, \{M^s\tilde{M}^k\varepsilon_0\}) = \varphi_M(\{M\tilde{M}^k\varepsilon_0\})\,.$$

Ist aber die Folge $\{M^k\varepsilon_0\}$ gleichverteilt, so ist auch die Folge $\{\tilde{M}^k\varepsilon_0\}$ gleichverteilt (vgl. Postnikow [2]). Nach dem Weylschen Kriterium ergibt sich für beliebiges natürliches $m > 0$

$$\sum_{k=0}^{N-1} \exp(2\pi i m\varepsilon_0\tilde{M}^k) \doteq o(N)\,;$$

insbesondere gilt diese Gleichung für alle $m = m_1 M \neq 0$. Das bedeutet aber, daß auch die Folge $\{M\tilde{M}^k\varepsilon_0\}$ gleichverteilt ist. Wie schon oben folgt daraus

$$\lim_{M\to\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} f(\tilde{\boldsymbol{x}}_k) = J\,,$$

und der Satz ist bewiesen.

Die folgenden beiden Sätze, die vom Autor gefunden wurden, erlauben es, eine Aussage über die Grenzverteilung des arithmetischen Mittels

$$\frac{1}{N}\sum_{k=0}^{N-1} f(\tilde{\boldsymbol{x}}_k)$$

für $M \to \infty$ und $N \to \infty$ zu machen.

Lemma 7.3. *Ist $f(\{M^k x\}, \ldots, \{M^{k+s-1}x\})$, $k \geqq 0$, im Riemannschen Sinn über* (0, 1) *integrierbar, so gilt die Beziehung*

$$\int_0^1 f(\{M^k x\}, \ldots, \{M^{k+s-1}x\})\,dx = \int_0^1 f(x, \{Mx\}, \ldots, \{M^{s-1}x\})\,dx\,.$$

Zum Beweis genügt es, (0, 1) in M^k gleiche Intervalle zu zerlegen und dann wie beim Beweis von Satz 7.10 vorzugehen.

Wir führen nun die Bezeichnung

$$J_M[m_1, \ldots, m_k] = \int_0^1 f^{m_1}(\{M^{i_1}x\}, \ldots, \{M^{i_1+s-1}x\}) \cdots f^{m_k}(\{M^{i_k}x\}, \ldots, \{M^{i_k+s-1}x\})\,dx$$

ein. Dabei sind $m_1, \ldots, m_k$, $i_1, \ldots, i_k$ natürliche Zahlen mit $i_p + s \leqq i_{p+1}$, $p = 1, \ldots, k-1$, und die Existenz des Integrals wird vorausgesetzt. (Die Funktion f soll außerdem in $\mathscr{D}_s$ beschränkt sein.)

Lemma 7.4. *Bei beliebiger Wahl der natürlichen Zahlen $m_1, \ldots, m_k$ gilt die Beziehung*

$$\lim_{M\to\infty} J_M[m_1, \ldots, m_k] = \int_{\mathcal{D}_s} f^{m_1}(\boldsymbol{x})\, d\boldsymbol{x} \cdots \int_{\mathcal{D}_s} f^{m_k}(\boldsymbol{x})\, d\boldsymbol{x}\,.$$

Beweis. Nach Lemma 7.3 kann man die Potenz von M in dem Ausdruck für $J_M[m_1, \ldots, m_k]$ reduzieren. Der weitere Beweis ergibt sich dann aus Satz 7.10, wenn man berücksichtigt, daß der Integrand eine Funktion φ_M ist, die einer Funktion von ks Veränderlichen

$$f^{m_1}(x_1^{(1)}, \ldots, x_s^{(1)}) \cdots f^{m_k}(x_1^{(k)}, \ldots, x_s^{(k)})$$

entspricht. Wir betrachten nun die Differenz

$$\frac{1}{N}\sum_{k=0}^{N-1} f(\tilde{\boldsymbol{x}}_k) - J = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{f}(\tilde{\boldsymbol{x}}_k)$$

mit $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - J$. Nach Lemma 7.4 gilt

$$\lim_{M\to\infty} \int_0^1 \left[\frac{1}{N}\sum_{k=0}^{N-1} f(\tilde{\boldsymbol{x}}_k) - J\right]^p d\varepsilon_0 = N^{-p} \int_{\mathcal{D}_s} d\boldsymbol{x}_1 \cdots \int_{\mathcal{D}_s} d\boldsymbol{x}_N \left[\sum_{k=0}^{N-1} \tilde{f}(\boldsymbol{x}_k)\right]^p \equiv I_p[N]\,.$$

Ist $\boldsymbol{\alpha}$ eine in $\mathcal{D}_s$ gleichverteilte Zufallsgröße und wird $\xi = f(\boldsymbol{\alpha})$ gesetzt, so ist $I_p[N]$ das p-te Moment des Stichprobenmittels von ξ. Daraus ergibt sich

$$\lim_{M\to\infty} \operatorname{mes} \left\{\varepsilon_0 : \frac{1}{N}\sum_{k=0}^{N-1} \tilde{f}(\tilde{\boldsymbol{x}}_k) < y\right\} = F_N(y)\,, \tag{7.2.4}$$

wobei $F_N(y)$ die Verteilungsfunktion des Stichprobenmittels von ξ ist. Für hinreichend große N nähert sich diese Verteilung einer Normalverteilung mit dem Erwartungswert Null und der Dispersion

$$\sigma_N^2 = N^{-1}\left[\int_{\mathcal{D}_s} (f(\boldsymbol{x}) - J)^2\, d\boldsymbol{x}\right].$$

Es ist leicht zu sehen, daß für den Mittelwert $N^{-1}\sum_{k=0}^{N-1} f(\boldsymbol{x}_k)$ (die $\boldsymbol{x}_k$ werden aus der Folge $\mathcal{E} = \varepsilon_0, \varepsilon_1, \ldots$ nach der Vorschrift $\boldsymbol{x}_k = (\varepsilon_{k+1}, \ldots, \varepsilon_{k+s})$ gebildet) ein anderes Resultat gilt. Die entsprechenden Zufallsgrößen ξ sind dann s-abhängig.

Die im ersten Kapitel beschriebenen linearen Rekursionsverfahren zur Erzeugung von Pseudozufallszahlen können in gewissem Maß durch diese Resultate begründet werden. Man kann hoffen, für diese Verfahren solche Parameter zu finden, daß sich eine asymptotische Normalverteilung ergibt. Es ist klar (und das wurde auch mehrfach in der Literatur behauptet), daß man dann gute Resultate bei der Berechnung von Pseudozufallszahlen mit einer großen Anzahl gültiger Ziffern erwarten kann. Diese Frage bedarf aber noch weiterer Untersuchungen. Anschließend gehen wir kurz auf Resultate ein, die sich auf Pseudozufallszahlenfolgen beziehen, wenn die Berechnung mit einer festen endlichen Ziffernzahl durchgeführt wird.

2. Folgen von Zahlen mit endlich vielen Ziffern. Ihre Periodizität. Die Folge der Gestalt (7.2.3) besitzt bei einem festen M unter der Voraussetzung, daß ihre Glieder mit endlicher Ziffernzahl berechnet werden, die wichtige Eigenschaft eines außerordentlich einfachen Rechenverfahrens. Aus diesem Grund wird die genannte Folge recht oft bei der Lösung praktischer Aufgaben mit der Monte-Carlo-Methode benutzt.

Wie sich aus den früher beschriebenen Beispielen konkreter Aufgaben ergibt, bildet die Einfachheit des Algorithmus zur Erzeugung von Pseudozufallszahlen eine nicht unwichtige Eigenschaft. Zugleich fehlen aber Resultate, die eine Beschreibung der Funktionenklasse gestatten, für die die näherungsweise Integration mit Hilfe solcher Folgen zu einem konvergenten Verfahren führt. Obwohl eine Reihe von Folgen durch statistische Teste geprüft wurde (ausführliche Angaben darüber folgen in diesem Abschnitt), ist bei ihrer Anwendung eine gewisse Vorsicht notwendig, insbesondere in den Fällen, in denen die Rechenergebnisse nicht mit entsprechenden experimentellen Daten verglichen werden können.

Wir geben anschließend nur Resultate an, die sich auf die Periodenlänge der durch lineare Rekursionsformeln erzeugten Folgen beziehen. Nur kurz gehen wir auf einige andere Untersuchungen ein, die diese Verfahren betreffen.

Die nun folgenden Resultate können auf verschiedene Weise aus den Ergebnissen der elementaren Zahlentheorie hergeleitet werden. Eine der einfachsten Darstellungen findet sich in der Monographie von JANSSON [1], der wir im wesentlichen folgen.

Wir kommen nun auf die Verfahren zur Erzeugung von Pseudozufallszahlen zurück, über die wir kurz im ersten Kapitel berichteten. Wir betrachten eine Folge $X = x_0, x_1, \ldots$ ganzer Zahlen, die durch die Beziehung

$$x_{n+1} \equiv a_0 x_n + a_1 x_{n-1} + \cdots + a_k x_{n-k} + b \pmod P \tag{7.2.5}$$

mit nichtnegativen ganzen Zahlen $P, a_0, \ldots, a_k, b, x_0, \ldots, x_k$ definiert ist. Die Relation $x \equiv y \pmod P$ bedeutet in diesem Fall, daß x der Rest ist, der bei der Division von y durch P entsteht. Daher sind die Glieder der Folge X nicht größer als P, und die Zahlen x_i/P, $i = k + 1, k + 2, \ldots$, werden gewöhnlich als Pseudozufallszahlen auf dem Intervall $(0, 1)$ benutzt. Für $k = 0$ ist unschwer die Analogie zwischen der Folge (7.2.5) und der im vorigen Abschnitt betrachteten Folge festzustellen.

Wir werden nun voraussetzen, daß die größten gemeinsamen Teiler $(x_0, \ldots, x_k, b, P)$ und $(a_0, \ldots, a_k, P)$ der Zahlen $x_0, \ldots, x_k, b, P$ bzw. $a_0, \ldots, a_k, P$ gleich Eins sind Um die Abhängigkeit der Resultate von P hervorzuheben, werden wir die Folge X selbst mit $X(P)$ und ihre Periodenlänge mit $L(P)$ bezeichnen. Sind die Primfaktoren von P kleine Zahlen, so kann das Verfahren zur Bestimmung von $L(P)$ leicht auf ein einfaches Verfahren zur Bestimmung der Periodenlänge dieser kleinen Faktoren zurückgeführt werden. Zu diesem Zweck beweisen wir bei festen Werten $x_0, \ldots, x_k$ folgendes Lemma:

Lemma 7.5. *Es seien m_1 und m_2 positive ganze Zahlen, $(m_1, m_2) = 1$ und $L(m_1)$ und $L(m_2)$ die Periodenlänge von $X(m_1)$ bzw. $X(m_2)$. Dann ist $L(m_1 m_2) =$* k.g.V. *$(L(m_1), L(m_2))$, wobei* k.g.V. *das kleinste gemeinsame Vielfache der darauffolgenden Zahlen bedeutet.*

Beweis. Wird

$$y_{n+1} = a_0 y_n + \cdots + a_k y_{n-k} + b\,, \quad n \geqq k\,, \quad y_n = x_n\,, \quad 0 \leqq n \leqq k\,, \tag{7.2.6}$$

gesetzt, dann gilt nach den Voraussetzungen des Lemmas

$$y_{i+jL(m_1)} = x_{i+c_{i,j}m_1}\,, \tag{7.2.7}$$

$$y_{i+lL(m_2)} = x_{i+c_{i,l}m_2} \tag{7.2.8}$$

für $i = 0, 1, \ldots, k$, wobei $c_{i,l}$ und $c_{i,j}$ ganze und j und l natürliche Zahlen sind. Außerdem gilt

$$y_{i+L(m_1m_2)} = x_{i+c_im_1m_2} \tag{7.2.9}$$

mit einer ganzen Zahl c_i. $L(m_1m_2)$ ist die kleinste ganze Zahl, für die die Beziehungen (7.2.7) bis (7.2.9) erfüllt sind, also k.g.V. $(L(m_1), L(m_2))$, womit das Lemma bewiesen ist.

Aus dem Lemma ergibt sich leicht eine Folgerung

Folgerung. *Ist* $P = \sum_{i=1}^{s} p_i^{n_i}$ *die Primfaktorenzerlegung der Zahl* P, *dann ist*

$$L(P) = \text{k.g.V.}\left(L(p_1^{n_1}), \ldots, L(p_s^{n_s})\right).$$

Der für uns wichtige Fall mehrfacher Faktoren wird in den folgenden beiden Lemmata untersucht.

Lemma 7.6. *Wenn die Folge* $X(p^m)$ *die Periodenlänge* $L(p^m)$ *besitzt* (p *Primzahl,* m *natürliche Zahl*), *dann ist die Periodenlänge der Folge* $X(p^{m+1})$ *entweder auch gleich* $L(p^m)$ *oder aber* p*-mal größer.*

Beweis. y_{n+1} ist wie oben durch (7.2.6) definiert. Dann gilt

$$y_{i+jL(p^m)} = x_{i+c_{i,j}p^m}\,, \qquad i = 0, \ldots, k\,,$$

wobei $c_{i,j}$ nichtnegativ ist. Ist $c_{i,1}$ für alle i durch p teilbar ist, dann gilt $L(p^{m+1}) = L(p^m)$. Andernfalls gibt es eine natürliche Zahl r, für die $c_{i,1} \equiv r \pmod p$ und folglich $c_{i,j} \equiv rj \pmod p$ gilt. Weil p eine Primzahl ist, gilt $jr \not\equiv 0 \pmod p$ für $j = 1, 2, \ldots, p-1$ und $jr \equiv 0 \pmod p$ für $j = p$. Daraus ergibt sich $L(p^{m+1}) = pL(p^m)$, womit der Beweis beendet ist.

Lemma 7.7. *Gilt für die Periodenlänge* $L(p^{m+1}) = pL(p^m)$, *so folgt daraus* $L(p^{l+m}) = p^l L(p^m)$ *mit natürlichen Zahlen* l *und* m.

Beweis. Für $l = 2$ erhält man unter Beibehaltung der vorherigen Bezeichnungen $c_{i,p} \equiv pr \pmod{p^r}$. Folglich ist $c_{i,jp} \equiv jpr \pmod p)^r$. Daraus folgt wie beim Beweis von Lemma 7.6 $jpr \equiv 0 \pmod{p^r}$, $j = 1, 2, \ldots, p-1$, und gleichzeitig $ppr \equiv 0 \pmod{p^r}$. Der Beweis für $l > 2$ kann leicht mit vollständiger Induktion geführt werden.

Wir gehen nun auf den wichtigsten Spezialfall, auf die Folge

$$x_{n+1} \equiv Mx_n \pmod{2^m} \tag{7.2.10}$$

ein. Die Untersuchung dieser Folgen für $m = 1, 2, 3, 4$ und verschiedene Werte für M liefert folgendes Resultat:

m	M	Zusammensetzung der Folge	Perioden-länge
1	1	1, 1, 1, . . .	1
2	1	1, 1, 1, . . .	1
	3	1, 3, 1, . . .	2
3	1	1, 1, 1, . . .	1
	3	1, 3 oder 5, 7	2
	5	1, 5 oder 3, 7	2
	7	1, 7 oder 3,5	2
4	1	1, 1, 1, . . .	1
	3	1, 3, 9, 11 oder 5, 15, 13, 7	4
	5	1, 5, 9, 13 oder 3, 15, 11, 7	4
	7	1, 7 oder 3, 5 oder 9, 15 oder 11, 13	2
	9	1, 9 oder 3, 11 oder 5, 13 oder 7, 15	2
	11	1, 11, 9, 3 oder 5, 7, 13, 15	4
	13	1, 13, 9, 5 oder 3, 7, 11, 15	4
	15	1, 15 oder 3, 13 oder 5, 11 oder 7, 9	2

Tabelle 7.1

Daraus kann man auf Grund der eben bewiesenen Lemmata leicht folgende Behauptung ableiten: Die maximale Periodenlänge der Folge (7.2.10) ergibt sich für $m \geqq 3$ zu 2^{m-2}. Diese Länge wird für $M \equiv 3 \pmod 8$ oder $M \equiv 5 \pmod 8$ und für alle ungeraden x_0 erreicht. In analoger Weise kann man auch die Folge

$$x_{n+1} \equiv Mx_n + b \pmod{2^m} \tag{7.2.11}$$

untersuchen, wobei sich folgendes Resultat ergibt: Die maximale Periodenlänge der Folge (7.2.11) beträgt 2^m und hängt nicht von x_0 ab. Diese Länge wird für $M \equiv 1 \pmod 4$ und $b \equiv 1 \pmod 2$ erreicht. Für Rechnungen auf Anlagen mit Dualzahlenarithmetik sind die angegebenen Resultate sehr interessant. Von Interesse sind auch Aussagen über die Zusammensetzung der entsprechenden Folgen im Fall maximaler Periodenlänge. Ohne Beweis geben wir die entsprechenden Resultate an.

Für die Folge (7.2.10) gilt:

1. Wenn $M \equiv 3 \pmod 8$ und $x_0 \equiv 1, 3, 9$ oder $11 \pmod{16}$ ist, dann besteht die Folge aus den Zahlen $8k + 1$ und $8k + 3$ für $k = 0, 1, ..., 2^{m-3} - 1$.

2. Wenn $M \equiv 3 \pmod 8$ und $x_0 \equiv 5, 7, 13$ oder $15 \pmod{16}$ ist, dann besteht die Folge aus den Zahlen $8k + 5$ und $8k + 7$ für $k = 0, 1, ..., 2^{m-3} - 1$.

3. Wenn $M \equiv 5 \pmod 8$ und $x_0 \equiv 1 \pmod 4$ ist, dann besteht die Folge aus den Zahlen $4k + 1$ für $k = 0, 1, ..., 2^{m-2} - 1$.

4. Wenn $M \equiv 5 \pmod 8$ und $x_0 \equiv 3 \pmod 4$ ist, dann besteht die Folge aus den Zahlen $4k + 3$ für $k = 0, 1, ..., 2^{m-2} - 1$.

Die Folge (7.2.11) besteht natürlicherweise (im Fall maximaler Periodenlänge) aus allen Zahlen von 0 bis $2^m - 1$.

Es gibt eine große Anzahl von Arbeiten, in denen statistische Teste zur Untersuchung der Folgen benutzt werden, die durch lineare Rekursionsformeln für verschiedene Parameterwerte erzeugt werden. Am vollständigsten wurden die Folgen der Gestalt (7.2.10) untersucht. Wir beschränken uns hier nur auf Hinweise auf einige Literaturstellen (hauptsächlich nichtsowjetische), wo solche Untersuchungen durchgeführt wurden oder die Benutzung linearer Rekursionsformeln beschrieben wurde. Diese bibliographischen Hinweise sind in Tabelle 7.2 (nach verschiedenen Werten für M und P geordnet) zusammengestellt. Eine weitere Bibliographie hierzu kann man bei JANSSON [1] finden. Beim praktischen Rechnen ist einer der am meisten gebrauchten Werte $M = 5^{15}$ mit $P = 2^{35}$ oder $P = 2^{36}$. Die Autoren heben für diesen M-Wert gewöhnlich die zusätzliche Erfüllung verschiedener statistischer Teste oder die gute Übereinstimmung mit dem Experiment hervor. Aussagen über Folgen der Gestalt (7.2.11) mit $b \neq 0$ kann man auch in der in Tabelle 7.2 genannten Literatur finden.

M	P	Autor
5^{13}	2^{39}	JUNCOSA [1]
5^{17}	2^{42}	TAUSSKY und TODD [1]
23	$2^{35} + 1$	JOHNSON [1]
3^{19}	10^{20}	BOFINGER [1]
23	$2^{31} - 1$	ESMENJAUD-BONNARDEL [1]
23	$10^8 + 1$	MACLAREN und MARSAGLIA [1]
$2^{17} + 3$	2^{35}	COVEYOU und MACPHERSON [1]
$10^5 + 3$	10^{10}	COVEYOU und MACPHERSON [1]
5^{15}	2^{35}	COVEYOU und MACPHERSON [1]
6469693231	2^{36}	ERMAKOW und PROKOFJEWA [1]
5^{17}	2^{40}	MARTSCHUK u. a. [1]
1050005	2^{35}	GELDER [1]
1000003		
551003		
551023		
400003		
200003		
20405		
20005		
20003		
10405		
10005		
10003		
6263		
4005		
4003		
2003		
203		

Tabelle 7.2

Analytische Untersuchungen, die mit den „pseudostatistischen Eigenschaften" der Folgen verbunden sind, werden beispielsweise in den Arbeiten von GOLENKO [1], JANSSON [1], COVEYOU und MACPHERSON [1] durchgeführt. Unter pseudostatistischen Eigenschaften einer Folge versteht man Eigenschaften integraler Art, die über die volle Periode berechnet werden. Solche Eigenschaften entsprechen gewöhnlich den Momenten einer Zufallsgröße. Wenn L die Periodenlänge der Folge ist und die Folge alle Zahlen $1, \ldots, L$ enthält, dann ist das Analogon des Erwartungswertes

$$\tilde{\mathsf{M}}x = \frac{1}{L} \sum_{i=1}^{L} x_i = \frac{1}{L} \sum_{i=1}^{L} i = \frac{L}{2} - \frac{1}{2}.$$

Entsprechend setzt man

$$\tilde{\mathsf{D}}x = \tilde{\mathsf{M}}x^2 - (\tilde{\mathsf{M}}x)^2 = \frac{L^2}{12} - \frac{1}{12}.$$

Was die Kovarianz zwischen den n-ten und $(n + k)$-ten Gliedern der Folge betrifft, so hängt der entsprechende Ausdruck von der Gestalt der Formel ab, nach der die Folge berechnet wurde, also insbesondere von den Konstanten M und b. Die Herleitung dieser Kovarianzformeln ist eine recht schwierige Aufgabe. Für die Folgen (7.2.10) und (7.2.11) wurden analytische Methoden zur Berechnung der Kovarianz $\operatorname{cov}(x_n, x_{n+k})$ von JANSSON [1] ausgearbeitet. Auf einer EDVA untersuchte er die empirische Verteilung der Korrelationskoeffizienten $\tilde{\varrho}_k$, die in den meisten Fällen genügend gut mit der theoretischen Verteilung übereinstimmte. Weil $\tilde{\varrho}_k$ von den Konstanten M und b abhängt und die Berechnung von $\tilde{\varrho}_k$ auf der Grundlage theoretischer Formeln erfolgen kann, können solche Berechnungen zu einer sinnvollen Auswahl dieser Konstanten benutzt werden. Ein anderer Weg, der auch von den Mittelwerten über eine vollständige Periode der Folge ausgeht, wurde von COVEYOU und MACPHERSON [1] vorgeschlagen. Dort wird die Übereinstimmung der Verteilung von nach Formel (7.2.5) erzeugten Folgen mit Hilfe einer sogenannten Wellenzahl charakterisiert.

Es sei z eine Veränderliche, die die ganzzahligen Werte $0, \ldots, L - 1$ annimmt. Dann gilt $z/L \in [0, 1)$. Wir führen nun das diskrete Maß $\mu_0(dz)$ ein, das in jedem Punkt z/L den Wert $1/L$ hat und erhalten so folgendes Lemma.

Lemma 7.8. *Die Funktion $f(x_1, \ldots, x_s)$ sei in den Punkten $(z_1/L, \ldots, z_s/L)$ definiert, wobei $z_1, \ldots, z_s$ ganzzahlige Werte zwischen 0 und $L - 1$, $L > 0$, sind. Dann ist in jedem der genannten Punkte die Gleichung*

$$f\left(\frac{z_1}{L}, \ldots, \frac{z_s}{L}\right) = \sum_{m_1, \ldots, m_s = 0}^{L-1} c_L(\boldsymbol{m}) \exp\left(2\pi i \frac{(\boldsymbol{m}, \boldsymbol{z})}{L}\right) \tag{7.2.12}$$

erfüllt (Zerlegung in eine endliche Fourierreihe). Dabei sind die

$$c_L(\boldsymbol{m}) = \frac{1}{L^s} \sum_{k_1, \ldots, k_s}^{L-1} f\left(\frac{k_1}{L}, \ldots, \frac{k_s}{L}\right) \exp\left(-2\pi i \frac{(\boldsymbol{m}, \boldsymbol{k})}{L}\right)$$

die endlichen Fourierkoeffizienten der Funktion f.

Der Beweis folgt aus der Orthogonalität des Funktionensystems $\{\exp(2\pi i k z)\}$ bezüglich des Maßes $\mu_0(dz)$. Die Gleichverteilung bezüglich des Maßes $\mu_0(dz)$ (mit der Dichte 1) ist natürlicherweise als Näherung für die Gleichverteilung auf $[0, 1)$ zu betrachten, als ihr diskretes Analogon. Außerdem muß 1 die Dichte der gemeinsamen Verteilung von Paaren, Tripeln usw. bis zu l-Tupeln ($l < L$) bezüglich des Produktes einer entsprechenden Anzahl von Maßen μ_0 sein.

Es seien $\varphi_1(z_1), \varphi_2(z_1, z_2), \ldots, \varphi_l(z_1, \ldots, z_l)$ die empirischen Dichten der gemeinsamen Verteilungen bezüglich des Maßes $(\mu_0)_l$, die einem konkreten Algorithmus zur Erzeugung von Pseudozufallszahlen entsprechen. Bei der Zerlegung dieser Dichten in eine Fourierreihe nach (7.2.12) werden nicht alle höheren Fourierkoeffizienten Null, sondern einige davon werden Eins sein. Es seien $\tilde{\boldsymbol{m}} = (\tilde{m}_1, \ldots, \tilde{m}_s)$ die Werte $\boldsymbol{m}$, für die die zugehörigen Fourierkoeffizienten Eins sind. Man setzt nun

$$\lambda = \frac{1}{|\tilde{\boldsymbol{m}}|} = (\tilde{m}_1^2 + \cdots + \tilde{m}_s^2)^{-1/2}$$

und nennt die Größe λ die Wellenzahl. Die Werte dieser Zahl können die Güte des Algorithmus zur Erzeugung von Pseudozufallszahlen charakterisieren. Offensichtlich darf eine gute Folge keine kleine Wellenzahl haben. Coveyou und Macpherson [1] haben die minimalen Wellenzahlen für einige Algorithmen analytisch oder auf einer EDVA berechnet. Dort werden auch einige Abschätzungen der minimalen Wellenzahl nach unten angegeben. Eine harmonische Analyse dieser Art ist jedenfalls leichter durchführbar als die unmittelbare Untersuchung der gemeinsamen Verteilungen von Pseudozufallszahlen.

Aus dieser kurzen Darstellung einiger Resultate über lineare Rekursionsformeln kann man schlußfolgern, daß ihre Anwendung in der Monte-Carlo-Methode im Grunde genommen auf einem „experimentellen" Vorgehen beruht (vgl. Kap. II, § 1).

§ 3. Die Benutzung der Stützstellen von Quadraturformeln in der Monte-Carlo-Methode

Wie in § 1 gezeigt wurde, ist die Benutzung einer vollständig gleichverteilten Folge als Folge von Realisierungen der Zufallsgröße α praktisch in allen Fällen sinnvoll. Freilich muß die Einschränkung gemacht werden, daß die Geschwindigkeit der Konvergenz des arithmetischen Mittels gegen das entsprechende Integral im allgemeinen von der Konstruktion der Folge und den Eigenschaften des Integranden abhängt. Die in § 1 beschriebene theoretische Konstruktion ist für praktische Rechnungen allerdings nicht verwendbar.

Die in § 2 ausführlich beschriebenen linearen Rekursionsverfahren haben eine gewisse Beziehung zu den vollständig gleichverteilten Folgen. Die mit diesen Verfahren erzeugten Folgen $x_0, x_1, \ldots$ mit endlicher Ziffernzahl werden sehr oft zur Lösung praktischer Aufgaben benutzt. Leicht ist aber folgendes zu sehen. Wenn die Größe $x_0(m_1 + m_2 M + \cdots + m_s M^{s-1})/P$ eine ganze Zahl ist, dann führt die Integration der Funktion $\exp\left(2\pi i(m_1 x_1 + \cdots + m_s x_s)\right)$ mit Hilfe solcher Folgen

zu keinem Erfolg. Der genannte Zusammenhang der linearen Rekursionsverfahren mit den vollständig gleichverteilten Folgen hat offensichtlich nur qualitativen Charakter, und in dieser Richtung sind noch weitere Untersuchungen notwendig. Das Problem, das hier entsteht, besteht in folgendem. Es gibt eine Methode der näherungsweisen Integration, die sehr einfach auf einer EDVA zu realisieren ist. Dann entsteht die Forderung, eine möglichst umfangreiche Funktionenklasse anzugeben, für die diese Integrationsmethode brauchbar ist. Der Begriff der Brauchbarkeit muß freilich erst in entsprechender Weise formalisiert werden.

Neben den durch lineare Rekursionsverfahren erzeugten Folgen werden bei der Anwendung der Monte-Carlo-Methode praktisch auch einige andere Folgen benutzt, deren Verhalten etwas ausführlicher untersucht wurde. Wir gehen nun sehr kurz auf Fragen ein, die mit der Benutzung dieser Folgen im Zusammenhang stehen, wobei wir im wesentlichen algorithmische Aspekte im Auge haben.

Wie schon am Schluß von Kap. IV, § 1, bemerkt wurde, hängen die statistischen Teste, die auf eine bestimmte Zahlenfolge formal angewendet werden, eng zusammen mit dem Fehler bei der Integration für eine hinreichend umfangreiche Funktionenklasse. Für das Analogon der Kolmogoroffschen Restgröße in bezug auf determinierte Folgen wird auch der Terminus „Diskrepanz" benutzt.

Ist $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ eine Punktfolge im s-dimensionalen Würfel

$$\mathcal{D}_s = \{(0 \leqq x_i < 1); \quad i = 1, \ldots, s\} ,$$

und $D_s(\boldsymbol{y})$ ein in diesem Würfel enthaltenes s-dimensionales Parallelepiped

$$D_s(\boldsymbol{y}) = \{0 \leqq x_i < y_i < 1; \quad i = 1, \ldots, s\} , \qquad \boldsymbol{y} = (y_1, \ldots, y_s) ,$$

dann heißt die Funktion $g(\boldsymbol{y}) = N^{-1}\nu(\boldsymbol{y})$ (N Anzahl der Glieder der Folge; $\nu(\boldsymbol{y})$ Anzahl derjenigen Glieder, die in $D_s(\boldsymbol{y})$ liegen) die *lokale Diskrepanz* der Folge aus N Punkten. Die Größe $\sup |g(\boldsymbol{z})| = D\{\boldsymbol{x}_n\}$, $\boldsymbol{z} \in \mathcal{D}_s$, die auch ein Analogon der Kolmogoroffschen Testgröße darstellt, heißt *Diskrepanz* der Folge. Man betrachtet aber auch andere Normen für die Funktion $g(\boldsymbol{z})$, beispielsweise

$$T\{\boldsymbol{x}_n\} = \left[\int_{\mathcal{D}_s} g^2(\boldsymbol{z})\, d\boldsymbol{z}\right]^{1/2}$$

(Analogon der Testgröße VON-MISES-SMIRNOW).

ROTH [1] erhielt für die Größe $T\{\boldsymbol{x}_n\}$ folgende Abschätzung nach unten:

$$N T\{\boldsymbol{x}_n\} \geqq \frac{(\log_2 N)^{(s-1)/2}}{4^{s+1}\big((s-1)!\big)^{1/2}} . \tag{7.3.1}$$

Da $T\{\boldsymbol{x}_n\} \leqq D\{\boldsymbol{x}_n\}$ gilt, ergibt sich daraus

$$D\{\boldsymbol{x}_n\} \geqq c_s N^{-1} (\log N)^{(s-1)/2} , \tag{7.3.2}$$

wobei die Konstante c_s nicht von N abhängt.

Es gibt eine Reihe von Arbeiten, in denen Folgen mit recht guten Diskrepanzeigenschaften konstruiert werden (z. B. HALTON [1], SOBOL [7]). Diese Folgen können bei Rechnungen mit der Monte-Carlo-Methode als Pseudozufallszahlen benutzt werden. In diesem Zusammenhang ist es richtig, noch einmal auf die schon zitierte Monographie von SOBOL [7] hinzuweisen. Dort werden diese Fragen recht ausführlich dargestellt, es werden verschiedene Kriterien für Gleichverteiltheit erörtert und eine umfangreiche Bibliographie angegeben. Auch die Arbeit von KOROBOW [4] enthält hierzu Aussagen.

Kann eine Aufgabe auf die Berechnung eines s-fachen Integrals über dem Einheitswürfel zurückgeführt werden, so kann im allgemeinen der klassische, in Kap. IV, § 1, genannte Weg Anwendung finden. Die einzige Besonderheit, die gewöhnlich im Zusammenhang mit der Lösung der in § 2 beschriebenen Aufgaben entsteht, ist die geringe Glattheit der zu integrierenden Funktionen. So verhält es sich aber nur, wenn die Vielfachheit s fest ist. Bei der Berechnung der Mittelwerte über die Trajektorien eines Zufallsprozesses ist es schwer, s im voraus zu bestimmen. Mehr noch, die Anzahl der Glieder in einer Trajektorie einer Markoffschen Kette mit Absorption stellt eine Zufallsgröße dar. In diesen Fällen muß der Charakter des Integrationsverfahrens sukzessiv sein, d. h., die neuen Stützstellen der Quadraturformel müssen zu den vorher benutzten ohne Änderung der letzten hinzugefügt werden. Es sind eine Reihe von Quadraturformeln hergeleitet worden, die dieser Forderung genügen. Sie fanden Anwendung bei der Lösung von Aufgaben des Strahlendurchgangs mit der Monte-Carlo-Methode (KOROBOW [3], S. 87–92; GELFAND, FEINBERG, FROLOW und TSCHENZOW [1], SOBOL [7]).

Ein anderer Weg besteht darin, daß bei der Integration der Funktion $f(x_1, ..., x_r, ..., x_s)$ nach den ersten r Veränderlichen eine Quadraturformel mit gleichen Gewichten gewählt wird. Die Veränderlichen $x_{r+1}, ..., x_s$ werden in jedem Glied der Quadratursumme durch die gleichen unabhängigen Realisierungen der gleichverteilten Zufallsgröße α ersetzt. Insbesondere kann für die Integration nach den ersten r Veränderlichen die Quadraturformel von KOROBOW (Kap. IV, § 1) benutzt werden. Für die Menge ihrer Stützstellen gilt übrigens folgende Abschätzung der Diskrepanz:

$$D_N\{\boldsymbol{x}_n\} \leq B(\log^\beta N)\, N^{-1}\,. \tag{7.3.3}$$

Dabei sind B und β von N unabhängige Konstanten. Die Abschätzung selbst folgt aus der Tatsache, daß die charakteristische Funktion des Parallelepipeds $D_s(\boldsymbol{y})$ zur Funktionenklasse E_s^1 gehört und aus der Abschätzung (4.1.16). Ein Vergleich der Abschätzungen (7.3.2) und (7.3.3) zeigt, daß die Methode der optimalen Koeffizienten zu einer Punktmenge führt, deren Diskrepanz sich von der minimalen nur um eine Größe von der Ordnung $\log^{\beta_1} N/N$ unterscheidet (β_1 ist eine von N unabhängige Konstante). Daher gibt es eine enge Beziehung zwischen der Monte-Carlo-Methode in der in den ersten sechs Kapiteln beschriebenen Form und den Quadraturformeln zur Integration von Funktionen geringer Glattheit. Es ist daran zu erinnern, daß die geringe Glattheit des Integranden mit der Art der Darstellung des Maßes, über das die Integration ausgeführt wird, verknüpft

ist. Methoden, mit deren Hilfe diese Darstellung des Maßes verwirklicht wird, wurden im zweiten Kapitel beschrieben.

Außerdem zeigt ein Vergleich der Umkehrformeln und der v.-Neumannschen Methode für die Modellierung von Zufallsgrößen, daß in vielen Fällen eine Vereinfachung des Algorithmus (Benutzung der v.-Neumannschen Methode oder einer ihrer Modifikationen anstelle der Umkehrformeln) zur einer Vergrößerung der Vielfachheit des zu berechnenden Integrals führt.

So erscheint vom Standpunkt der numerischen Mathematik die Vergrößerung der Vielfachheit des zu berechnenden Integrals und meistens auch die Verringerung der Glattheit des Integranden zugunsten einer Vereinfachung des Algorithmus als eine charakteristische Besonderheit der Monte-Carlo-Methode. Wie die Ergebnisse zeigen, ist ein solcher Weg, der durch die Entwicklung der Rechentechnik hervorgerufen wurde, in vielen Fällen sehr fruchtbar.

LITERATUR

ALBERT, G. E.

[1] A general theory of stochastic estimates of the Neumann series for the solution of certain Fredholm integral equations and related series, Sympos. on Monte Carlo Methods, ed. H. A. MEYER, J. Wiley & Sons, New York 1956, p. 37–46.

ANDERSSON, S. E.

[1] A Monte Carlo model for simulation of tank battles, Nord. Tidskr. information Behandl. **6** (1966), 89–100.

ANDREIEF, C.

[1] Note sur une relation entre les integrales définies des products des fonctions, Mem. Soc. Sci., Bordeaux **3** (1883), No 2, 1–14.

ANOLIK, M. W., und R. G. BARANZEW (Анолик, М. В. и Р. Г. Баранцев)

[1] Вторая итерация интегрального кинетического оператора в ударной волне одноатомного газа, Методы вычислений (сб.), Изд-во ЛГУ **5** (1968), 43—48.

BACHWALOW, N. S. (Бахвалов, Н. С.)

[1] О приближенном вычислении кратных интегралов, Вестник МГУ (1959), № 4, 3—18.

[2] Оценка в среднем остаточного члена квадратурных формул, Журнал вычислит. матем. и матем. физ. **1** (1961), 64—77.

[3] Об оптимальных оценках сходимости квадратурных процессов и методов интегрирования типа Монте-Карло на классах функций, Сб. „Числ. методы решения дифф. и интегр. уравнений", Наука, Москва 1964, стр. 5—63.

VON BAHR, B., and C. G. ESSEEN

[1] Inequalities for the r-th absolute moment of a sum of random variables $(1 \leqq r \leqq 2)$, Ann. Math. Stat. **36** (1965), 229–303.

BARASCHENKOW, W. S., W. M. MALZEW und W. D. TONEJEW (Барашенков, В. С., В. М. Мальцев и В. Д. Тонеев)

[1] Взаимодействие быстрых протонов с тяжелыми ядрами, Изв. АН СССР, сер. физ. **30** (1966), 322—327.

BELJAJEW, B. N., und I. W. ZARIZYNA (Беляев, Б. Н., и И. В. Царицына)

[1] Расчет взаимодействия быстрых нуклонов с ядрами методом случайных испытаний, Методы вычислений (сб.), Изд-во ЛГУ **1** (1963), 76—107.

BERNARDINI, G., E. T. BOOTH and S. T. LINDENBAUM

[1] The interactions of high energy nucleons with nuclei, Phys. Rev. **88** (1952), 1017–1026.

BLUM, J. R.

[1] Approximation methods which converge with probability one, Ann. Math. Stat. **25** (1954), 382–386.

BOFINGER, E., and V. J. BOFINGER

[1] On a periodic property of pseudorandom sequences, J. Assoc. Comp. Mach. **5** (1958), 261–265.

BOGATSCHEWA, A. A., W. A. PEREPUCHOW und S. E. RUCHMAN (Богачева, А. А., В. А. Перепухов и З. Е. Рухман)
[1] Применение метода Монте-Карло к расчету аэродинамических характеристик тел сложной формы в свободно-молекулярном потоке, Журнал вычислит. матем. и матем. физ. **8** (1968), 1395—1402.
BOLSCHEW, L. N. (Большев, Л. Н.)
[1] О преобразованиях случайных величин, Теория вероятностей и ее применения **4** (1959), 136—149.
BOTSCHEK, N. P. (Бочек, Н. П.)
[1] О вычислении коэффицинтов Ритца и Галеркина методом Монте-Карло, Журнал вычислит. матем. и матем. физ. **6** (1966), 172—177.
BOX, G. E. P., and N. R. DRAPER
[1] The choice of a second order rotatable design, Biometrika **60** (1963), 355—362.
BUSLENKO, N. P.
[1] Simulation von Produktionsprozessen, Teubner-Verlag, Leipzig 1971 (Übersetzung aus dem Russischen).
[2] Modellierung komplizierter Systeme, Verlag Die Wirtschaft, Berlin 1972 (Übersetzung aus dem Russischen).
BUSLENKO, N. P., D. I. GOLENKO, I. M. SOBOL, W. G. SRAGOWITSCH und J. A. SCHREIDER (Бусленко, Н. П., Д. И. Голенко, И. М. Соболь, В. Г. Срагович и Ю. А. Шрейдер)
[1] Метод статистических испытаний (метод Монте-Карло), СМБ, Физматгиз, Москва 1962.
BUSLENKO, N. P., und J. A. SCHREIDER
[1] Die Monto-Carlo-Methode und ihre Verwirklichung mit elektronischen Digitalrechnern, Teubner-Verlag, Leipzig 1964 (Übersetzung aus dem Russischen).
BUTLER, J. W.
[1] Machine sampling from given probability distributions, Sympos. on Monte Carlo Methods, ed. H. A. MEYER, J. Wiley & Sons, New York 1956, p. 249—264.

CHAMPERNOWNE, O. G.
[1] The construction of the decimals normal in the scale of ten, J. London Math. Soc. 8 (1933), 254—260.
CHISSAMUTDINOW, A. I. (Хисамутдинов, А. И.)
[1] Об эффективности метода математических ожиданий для задач одного класса, Журнал вычислит. матем. и матем. физ. **7** (1967), 946—953.
[2] Решение некоторых вопросов теории метода Монте-Карло и двух задач ядерной геофизики, Диссертация канд. физ.-мат. наук, Новосибирск 1968.
CHOROFAS, D. N.
[1] Systems and simulation, Academic Press, New York—London 1965.
CIGLER, J., und G. HELMBERG
[1] Neuere Entwicklungen der Theorie der Gleichverteilung, Jahresber. Dtsch. Math.-Ver. **64** (1961), 1—50.
COLEMAN, W. A.
[1] Mathematical verification of a certain Monte Carlo sampling technique and applications of the technique to radiation transport problems, Nuclear Sci. and Eng. **32** (1968), No 1.
COVEYOU, R. R., and R. P. MACPHERSON
[1] Fourier analysis of uniform random number generators, J. Assoc. Comp. Mach. **14** (1967), 100—119.
COX, D. R., and W. L. SMITH
[1] Queues, Methuen & Co., Ltd./J. Wiley & Sons., Inc., New York 1961.

DJADKIN, I. G., und S. A. SHUKOWA (Дядькин, И. Г., и С. А. Жукова)

[1] О постановке задачи и алгоритме решения уравнения Шредингера методом Монте-Карло, Журнал вычислит. матем. и матем. физ. **8** (1968), 222 до 229.

DOOB, J. L.

[1] Stochastic processes, J. Wiley & Sons, New York 1953.

DVORETZKY, A.

[1] On stochastic approximation, Proc. 3rd Berkeley Sympos. on the Math. Stat. and Prob., ed. J. NEYMAN, Univ. of California Press, Berkeley 1956, p. 39–55.

DYNKIN, E. B., und A. A. JUSCHKEWITSCH

[1] Sätze und Aufgaben über Markoffsche Prozesse, Springer-Verlag, Berlin–Heidelberg–New York 1969 (Übersetzung aus dem Russischen).

ELEPOW, B. S., und G. A. MICHAILOW (Елепов, Б. С., и Г. А. Михайлов)

[1] О решении задачи Дирихле для уравнения $\Delta u - cu = -g$ моделированием блужданий по сферам, Журнал вычислит. матем. и матем. физ. **9** (1969), 647—654.

ERMAKOW, S. M. (Ермаков, С. М.)

[1] Интерполирование по случайным точкам, Журнал вычислит. матем. и матем. физ. **3** (1963), I, 186—189.

[2] Случайные квадратуры повышенной точности, Журнал вычислит. матем. и матем. физ. **4** (1964), 3, 550—554.

[3] Об оценивании параметров при непрямых измерениях, Труды МИАН СССР **79** (1965), 106—112.

[4] Письмо в редакцию, Теория вероятностей и ее применения **11** (1966), 728.

[5] О допустимости процедур метода Монте-Карло, ДАН СССР **172** (1967), 262—263

ERMAKOW, S. M., B. A. EFIMENKO, W. G. SOLOTUCHIN, F. F. MICHAILUS, S. S. NIKOLAISCHWILI und W. P. POLIWANSKI (Ермаков, С. М., Б. А. Ефименко, В. Г. Золотухин, Ф. Ф. Михайлус, Ш. С. Николайшвили и В. П. Поливанский)

[1] Методы решения уравнения переноса в неоднородных и ограниченных средах, 28 (A Conf.) 363 IУ-1964, 3-я Международная конференция ООН по использованию атомной энергии в мирных целях.

ERMAKOW, S. M., und S. I. PROKOFJEWA (Ермаков, С. М., и З. И. Прокофьева)

[1] О расчете защиты методом Монте-Карло, Вопросы физики защиты реакторов (сб.), вып. 2, Атомиздат, Москва 1966, стр. 67—71.

ERMAKOW, S. M., und W. G. SOLOTUCHIN (Ермаков, С. М., и В. Г. Золотухин)

[1] Полиномиальные приближения и метод Монте-Карло, Теория вероятностей и ее применения **5** (1960), 473—475.

ESMENJAUD-BONNARDEL, M.

[1] Un procédé de génération de nombres „Pseudo-aléatoires" pour Cab 500, Révue Franc. de Traitement de l'Information-Chiffres **7** (1964), 185–197.

EVANS, O. H.

[1] Applied multiplex sampling, Technometrics **5** (1963), 341–359.

FADDEJEW, D. K., und W. N. FADDEJEWA

[1] Numerische Methoden der linearen Algebra, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1973 (Übersetzung aus dem Russischen).

FAN, K.

[1] A generalisation of the Alaoglu-Bourbaki-theorem and its applications, Math. Z. **88** (1965), 48–60.

FANO, U., L. V. SPENCER and M. J. BERGER
[1] Penetration and diffusion of X-rays, Handbuch der Physik, Bd. 38, 2. Teil, Springer-Verlag, Berlin—Göttingen—Heidelberg 1959, S. 660—817.
FELLER, W.
[1] An introduction to probability theory and its applications, Vol. 2, J. Wiley & Sons, New York—London Sidney 1966.
FINNEY, D. J.
[1] An introduction to the theory of experimental design, Chicago Univ. Press, Chicago—London 1960.
FISHER, M. E.
[1] The nature of critical points, Univ. of Colorado Press, Boulder 1965.
FLECK, J. A.
[1] The calculation of nonlinear radiation transport by a Monte Carlo Method, Methods Comp. Phys. **1** (1963), 43—65.
FORSYTHE, G. E., and R. Z. LEIBLER
[1] Matrix inversion by a Monte Carlo methods, Math. Tables Aids Comp. **4** (1950), 127—129.
FRANKLIN, J. N.
[1] Deterministic simulation of random processes, Math. Comp. **17** (1963), 28—59.
[2] Numerical simulation of stationary and nonstationary Gaussian random processes, SIAM Rev. **7** (1965), 68—80.
FROLOW, A. S. (Фролов, А. С.)
[1] Решение интегральных уравнений теории переноса методом статистических испытаний, Диссертация канд. физ.-мат. наук, МИАН СССР, Москва 1963.
FROLOW, A. S., und N. N. TSCHENZOW (Фролов, А. С., и Н. Н. Ченцов)
[1] О вычислении методом Монте-Карло определенных интегралов, зависящих от параметров, Журнал вычислит. матем. и матем. физ. **2** (1962), 714—717.
[2] Использование зависимых испытаний в методе Монте-Карло для получения гладких кривых, Труды VI Всесоюзного совещания по теории вероятностей и матем. статистике, Гос. изд-во полит. и науч. лит-ры Литовской ССР, Вильнюс 1962, стр. 425—437.
[3] Решение трех типичных задач теории переноса методом Монте-Карло, Метод Монте-Карло в проблеме пореноса излучений (сб.), Атомиздат, Москва 1967, стр. 25—52.

GARABEDJAN, S.M., und I.W.ROMANOWSKI (Гарабедян, С. М., и И. В. Романовский)
[1] Приближенное построение доверительных интервалов для параметров марковской цепи, Теория вероятностей и ее применения **12** (1967), 123—127.
GELDER, A.
[1] Some new results in pseudo-random number generation, J. Assoc. Comp. Mach. **14** (1967), 785—792.
GELFAND, I. M., S. M. FEINBERG, A. S. FROLOW und N. N. TSCHENZOW
(Гельфанд, И. М., С. М. Фейнберг, А. С. Фролов и Н. Н. Ченцов)
[1] О применении метода случайных испытаний (метода Монте-Карло) для решения кинетического уравнения, Труды 2-й Международной Женевской конференции по применению атомной энергии в мирных целях, т. 5, Женева 1958, стр. 32—45.
GELFAND, I.M., A.S. FROLOW und N.N. TSCHENZOW (Гельфанд, И.М., А.С. Фролов, и Н. Н. Ченцов)
[1] Вычисление континуальных интегралов методом Монте-Карло, Известия вузов, сер. матем. **6** (1958), № 5, 32—45.

Gladyschew, E. G. (Гладышев, Е. Г.)
[1] О стохастической аппроксимации, Теория вероятностей и ее применения **10** (1965), 297—300.
Gnedenko, B. W.
[1] Lehrbuch der Wahrscheinlichkeitsrechnung, 5. Aufl., Akademie-Verlag, Berlin 1968 (Übersetzung aus dem Russischen).
Gnedenko, B. W., und A. N. Kolmogorov
[1] Grenzverteilungen von Summen unabhängiger Zufallsgrößen, 2. Aufl., Akademie-Verlag, Berlin 1960 (Übersetzung aus dem Russischen).
Golenko, D. I. (Голенко, Д. И.)
[1] Моделирование и статистический анализ псевдослучайных чисел на електронных вычислительных машинах, Наука, Москва 1965.
Granowski, B. L. (Грановский, Б. Л.)
[1] Некоторые вопросы планирования регрессионных экспериментов и теории квадратурных формул со случайными узлами, Диссертация, Ленинград 1968.
[2] О случайных квадратурах гауссовского типа, Журнал вычислит. матем. и матем. физ. **8** (1968), 879—884.
[3] О допустимости случайных квадратур гауссовского типа, Журнал вычислит. матем. и матем. физ. **11** (1971), 45—50.
Granowski, B. L., und S. M. Ermakow (Грановский, Б. Л., и С. М. Ермаков)
[1] О непараметрическом подходе к задачам планирования регрессионных экспериментов, ДАН СССР **180** (1968), 273—275.
[2] Случайные квадратуры с частично фиксированными узлами, Методы вычислений (сб.), Изд-во ЛГУ **6** (1970), стр. 79—88.

Haji-Sheikh, A., and E. M. Sparrow
[1] The floating random walk and its application to Monte Carlo solutions of heats equations, SIAM J. Appl. Math. **14** (1966), 370—389.
Hall, A.
[1] On an experiment determination of π, Messeng. Math. **2** (1873), 113—114.
Halton, J. H.
[1] On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, Numerische Mathematik **2** (1960), 84—90 and corrigenda p. 196.
[2] Sequential Monte Carlo, Proc. Cambr. Philos. Soc. **58** (1962), 57—78.
[3] A retrospective and prospective survey of the Monte Carlo method, SIAM Rev. **12** (1970), 1—63.
Hammersley, J. M., and D. C. Handscomb
[1] Monte Carlo methods, Methuen & Co., London/J. Wiley & Sons, New York 1964.
Hammersley, J. M., and K. W. Morton
[1] A new Monte Carlo technique — antithetic variates, Proc. Cambr. Philos. Soc. **52** (1956), 449—475.
Handscomb, D. C.
[1] Remarks on a Monte Carlo integration method, Numer. Math. **6** (1964), 261—268.
Hannan, E. J.
[1] Time series analysis, J. Wiley & Sons, Inc., New York—London 1960.
Haviland, J. K., M. L. Lavin and L. Trilling
[1] Application on the Monte Carlo method to rarified gas flows, Fluid Dynamics Trans., Pergamon Press, **1** (1964), 279—286.

Hicks, B. L., and M. A. Smith
[1] On the accuracy of Monte Carlo solutions of the nonlinear Boltzmann equations, J. Comp. Phys. **3** (1968), 58–79.

Ibragimow, I. A., und J. W. Linnik (Ибрагимов, И. А., и Ю. В. Линник)
[1] Независимые и стационарно связанные величины, Наука, Москва 1965.
Ishii, T., T. Secine and K. Ono
[1] A Monte Carlo calculation of gamma-ray scattering on the jeidac computer, Codes for reactor Comp., Vienna I, Austria 1961, p. 53–66.

Jansson, B.
[1] Random number generators, Almquist & Wiksell, Stockholm 1966.
Johnson, D. L.
[1] Generating and testing pseudo-random numbers on the IBM type 701, Math. Tables Aids Comp. **10** (1956), 8–13.
Juncosa, M. L.
[1] Random number generation on the BRL high-speed computing machines, Ballistic Research Lab., Report 855 (1953), Aberdeen Prov. Ground.

Kabak, J. W.
[1] Stopping rules for queuing simulation, Operat. Rev. **16** (1968), 431–437.
Kac, M.
[1] Probability and related topics in physical sciences, Interscience Publishers, New York – London 1959.
[2] Kilka zagadnień stochastycznych fizyki i matematyki, PWN, Warszawa 1961 (russische Übersetzung 1967).
Kadyrow, M. (Кадыров, М.)
[1] Таблицы случайных чисел, Изд-во СГУ, Ташкент 1936.
Kahn, H.
[1] Use of different Monte Carlo sampling techniques, Sympos. on Monte Carlo methods, ed. H. A. Meyer, J. Wiley & Sons, New York 1956, p. 145–191.
Kalos, M. H.
[1] On the estimation of flux at a point by Monte Carlo, Nuclear Sci. and Eng. **16** (1963), 111–117.
[2] Monte Carlo integration of the adjoint gamma-ray transport equation, Nuclear Sci. and Eng. **33** (1968), 284–290.
Kantorowitsch, L. W., und G. P. Akilow
[1] Funktionalanalysis in normierten Räumen, Akademie-Verlag, Berlin 1960 (Übersetzung aus dem Russischen).
Kaufmann, A., et R. Cruon
[1] Les phénoménes d'attente (Théorie et applications), Dunod, Paris 1961.
Kemeny, J. G., and J. L. Snell
[1] Finite Markov chains, D. van Nostrand Co., Princeton (N. J.)–London 1960.
Kendall, M. G., and B. Babington-Smith
[1] Tables of random sampling numbers, Tracts for Comp., Cambr., **24** (1939).
Kertiss, D. (Кертисс, Д.)
[1] Методы Монте-Карло для итерации линейных операторов, Успехи матем. наук **12** (1957), № 5, 149—174.
Kiefer, J.
[1] Optimum designs for regression problem, Ann. Math. Stat. **32** (1961), 268–279.

KIEFER, J., and J. WOLFOWITZ
[1] Stochastic estimation of the maximum of a regression function, Ann. Math. Stat. **23** (1952), 463—466.
KLIMOW, G. P. (Климов, Г. П.)
[1] Стохастические системы обслуживания, Наука, Москва 1966.
KOKSMA, J. F., and R. SALEM
[1] Uniform distribution and Lebesgue integration, Acta Sci. Math. Szeged **12** (1950), 87—96.
KOLMOGOROFF, A. N.
[1] Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin 1933.
[2] On tables of random numbers, Sankhya, Ser. A., **25** (1963), 369—376.
KOROBOW, N. M. (Коробов, Н. М.)
[1] О некоторых вопросах равномерного распределения, Изв. АН СССР, сер. матем., **14** (1950), 215—231.
[2] Оценки тригонометрических сумм с вполне равномерно распределенными функциями, ДАН СССР **133** (1960), 1011—1014.
[3] Теоретико-числовые методы в приближенном анализе, Физматгиз, Москва 1963.
[4] О некоторых вопросах теории диофантовых приближений, Успехи матем. наук **22** (1967), № 3, 83—118.
KRONMAL, R.
[1] Evaluation of a pseudorandom normal number generator, J. Assoc. Comp. Mach. **11** (1964), 357—363.
KRYLOW, W. I. (Крылов, В. И.)
Приближенное вычисление интегралов, Физматгиз, Москва 1959.
KRYLOW, W. J. (Крылов, В. Ю.)
[1] Разностные схемы для уравнений параболического типа и континуальные интегралы (кн. С. К. Годунов и В. С. Рябенький, Введение в теорию разностных схем, Физматгиз, Москва 1962, стр. 330—340).
KUBILJUS, I. P., und J. W. LINNIK (Кубилюс, И. П., и Ю. В. Линник)
[1] Арифметическое моделирование броуновского движения, Изв. вузов **6** (1959), 88—89.
KUROPATENKO, E. S., und W. N. OGIBIN (Куропатенко, Э. С., и В. Н. Огибин)
[1] Об одной схеме моделирования траекторий частиц в системах сложной геометрии, Журнал вычислит. матем. и матем. физ. **8** (1968), 212—216.
KUTAI, A. K., und G. S. FAINSCHTEIN (Кутай, А. К., и Г. З. Файнштейн)
[1] Об оценке резко выделяющихся наблюдений, Измерительная техника **1** (1967), 17—19.

LEHMANN, E. L.
[1] Testing statistical hypotheses, J. Wiley & Sons, Inc., New York/Chapman & Hall, Ltd., London 1959.
LEHMER, P. H.
[1] Mathematical methods in large-scale computing units, Ann. Comp. Lab. Harvard Univ. **26** (1951), 141—146.
LEIPUNSKI, O. I., u. a. (Лейпунский, О. И., и др.)
[1] Распространение гамма-квантов в веществе, Физматгиз, Москва 1960.
LÉVY, P.
[1] Processus stochastiques et mouvement brownien, Gauthier-Villars, Paris 1948.
LINNIK, J. W.
[1] Methode der kleinsten Quadrate in moderner Darstellung, VEB Deutscher Verlag der Wissenschaften, Berlin 1961 (Übersetzung aus dem Russischen).

MacLaren, M. D., and G. Marsaglia
[1] Uniform random number generators, J. Assoc. Comp. Mach. **12** (1965), 83–89.
Maiorow, L. W., und A. D. Frank-Kamenezki (Майоров, Л. В., и А. Д. Франк-Каменецкий)
[1] Сравнительная эффективность различных оценок в методе Монте-Карло, Препринт ИАЭ им. И. В. Курчатова, Москва 1969.
Markowitz, H. M., B. Hausner and H. W. Karr
[1] Simscript: A simulation programming language, Prentice-Hall, Englewood Cliffs (N. J.) 1963.
Marshall, A. W.
[1] The use of multi-stage sampling schemes in Monte Carlo computations, Sympos. on Monte Carlo methods, ed. H. A. Meyer, J. Wiley & Sons, New York 1956, p. 123–140.
Martschuk, G. I., und S. M. Ermakow (Марчук, Г. И., и С. М. Ермаков)
[1] Метод Монте-Карло и методы вычислительной математики, Метод Монте-Карло в проблеме переноса излучений (сб.), Атомиздат, Москва 1967.
Martschuk, G. I., G. A. Michailow, M. A. Nasaralijew und. R. A. Darbinjan (Марчук, Г. И., Г. А. Михайлов, М. А. Назаралиев и Р. А. Дарбинян)
[1] Решение прямых и некоторых обратных задач атмосферной оптики методом Монте-Карло, Наука, Сибирское отд. Новосибирск 1968.
Metropolis, N., and S. Ulam
[1] The Monte Carlo method, J. Amer. Stat. Assoc. **44** (1949), No. 247, 335–341.
Michailow, G. A. (Михайлов, Г. А.)
[1] О моделировании случайных величин для одного класса законов распределения, Теория вероятностей и ее применения **10** (1965), 749—751.
[2] К вопросу о построении экономичных алгоритмов моделирования случайных величин, Журнал вычислит. матем. и матем. физ. **6** (1966), 1134—1136.
[3] О применении экспоненциального преобразования в расчетах прохождения частиц методом Монте-Карло, Метод Монте-Карло в проблеме переноса излучений (сб.), Атомиздат, Москва 1967, стр. 67—72.
[4] Об одном принципе оптимизации расчетов по методы Монте-Карло, Журнал вычислит. матем. и матем. физ. **8** (1968), 1085—1093.
[5] Использование асимптотических решений проблемы Милна в расчетах по методу Монте-Карло, Изв. вузов, Физика **10** (1968), 76—79.
[6] Об одном классе оценок по методу Монте-Карло, Теория вероятностей и ее применения **15** (1970), 142—144.
Morton, K. W.
[1] A generalisation of antithetic variate technique for evaluating integrals, J. Math. Phys. **36** (1957), 289–293.
Müller, M. E.
[1] Some continuous Monte Carlo methods for the Dirichlet problem, Ann. Math. Stat. **27** (1956), 569–583.
[2] A note on method for generating points uniformly on N-dimensional spheres, Comm. Assoc. Math. **2** (1959), No. 4, 19–20.

Natanson, I. P.
[1] Konstruktive Funktionentheorie, Akademie-Verlag, Berlin 1955 (Übersetzung aus dem Russischen).
von Neumann, J.
[1] Various techniques used in connection with random digits. Monte Carlo method, Nat. Bur. Stand. Math. Series **12** (1951), 36–38.

OGIBIN, W. N. (Огибин, В. Н.)
[1] О применении расщепления и „рулетки“ в расчетах переноса частиц методом Монте-Карло, Метод Монте-Карло в проблеме переноса излучений (сб.), Атомиздат, Москва 1967, стр. 72—82.

PALEY, R. E. A. C., and N. WIENER
[1] Fourier transformation in the complex domain, Amer. Math. Soc., New York **1934**.

PIRANASCHWILI, S. A. (Пиранашвили, З. А.)
[1] К вопросу моделирования одного класса нестационарных случайных процессов, Вопросы исследования операции (сб.), Мецниереба, Тбилиси 1966.

POSTNIKOW, A. G. (Постников, А. Г.)
[1] Арифметическое моделирование случайных процессов, Труды МИАН СССР **57** (1960).
[2] Эргодические вопросы теории сравнений и диофантовых приближений, Труды МИАН СССР **82** (1966).

PROCHOROW, J. W. (Прохоров, Ю. В.)
[1] Сходимость случайных процессов и предельные теоремы теории вероятностей, Теория вероятностей и ее применения **1** (1956), 177—238.

PROCHOROW, J. W., und J. A. ROSANOW (Прохоров, Ю. В., и Ю. А. Розанов)
[1] Теория вероятностей. Основные понятия, предельные теоремы, случайные процессы, СМБ, Наука, Москва 1967 (englische Übersetzung 1969).

PUGH, E. L.
[1] A gradient technique of adaptive Monte Carlo, SIAM Rev. **8** (1966), 346—355.

RASSTRIGIN, L. A. (Расстригин, Л. А.)
[1] Статистические методы поиска, Наука, Москва 1968.

ROBBINS, H., and S. MONRO
[1] A stochastice approximation method, Ann. Math. Stat. **22** (1951), 400—407.

ROMANOWSKI, I. W. (Романовский, И. В.)
[1] О вероятностных задачах, приводящих к динамическому программированию, Труды Всесоюзного совещания по теории вероятностей и математической статистике, Ереван 1960, стр. 206—209.
[2] О методах моделирования непрерывных случайных величин из величин с равномерным распределением, Методы вычислений (сб.), Изд-во ЛГУ **3** (1966), 113—121.

ROSENBERG, L.
[1] Bernstein polynomials and Monte Carlo integration, SIAM J. Numerical Analysis **4** (1967), 566—577.

ROTH, K. F.
[1] On irregularities of distribution, Mathematika **1** (1954), 73—79.

SACKS, J.
[1] Asymptotic distribution of stochastic approximation procedures, Ann. Math. Stat. **29** (1958), 397—398.

SARYMSAKOW, G. A. (Сарымсаков, Г. А.)
[1] Основы теории процессов Маркова, Гостехиздат, Москва 1954.

SCHALAJEWSKI, O. W. (Шалавеский, О. В.)
[1] Проблема распределения наблюдений в полиномиальной регрессии, Труды МИАН СССР **79** (1965), 132—149.

SCHAPIRO-PJATEZKI, I. I. (Шапиро-Пятецкйи, И. И.)
[1] О законах распределения дробных долей показательной функции, Изв. АН СССР, сер. матем., **15** (1951), 47—52.

SCHARYGIN, I. F. (Шарыгин, И. Ф.)

[1] Оценки снизу погрешности квадратурных формул на классах функций, Журнал вычислит. матем. и матем. физ. **3** (1963), 370—376.

SCHEFFÉ, H.

[1] The analysis of variance, J. Wiley & Sons, New York/Chapman & Hall, London 1959.

SCHILOW, G. E., und FAN DYK TINH (Шилов, Г. Е., и Фан Дык Тинь)

Интеграл, мера и производная на линейных пространствах, Наука, Москва 1967.

SCHUR, I.

[1] Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten, Math. Z. **1** (1918), 373—385.

SISOWA, A. F. (Сизова, А. Ф.)

[1] Оценка плотности делений в сферическом реакторе, Метод Монте-Карло в проблеме переноса излучений (сб.), Атомиздат, Москва 1967, стр. 228—231.

SOBOL, I. M. (Соболь, И. М.)

[1] Псевдослучайные числа для машины ,,Стрела'', Теория вероятностей и ее применения **3** (1958), 205—211.

[2] Точная оценка погрешности многомерных квадратурных формул для функций классов W_1 и H_1, Журнал вычислит. матем. и матем. физ. **1** (1961), 208—216.

[3] О вычислении бесконечномерных интегралов, Журнал вычислит. матем. и матем. физ. **1** (1961), 918—922.

[4] Применение распределения ω^2 для оценки погрешности при вычислении интегралов методом Монте-Карло, Журнал вычислит. матем. и матем. физ. **2** (1962), 717—723.

[5] О периодах псевдослучайных последовательностей, Теория вероятностей и ее применения **9** (1964), 367—373.

[6] Детерминистическая интерпретация критериев согласия и проверка псевдослучайных чисел, Препринт, ИПМ АН СССР, Москва 1968. [1969]

[7] Многомерные квадратурные формулы и функции Хаара, Наука, Москва

[8] Об одном подходе к вычислению многомерных интегралов, Вопросы вычислительной и прикладной математики (сб.), Ташкент **38** (1970), 100—111.

SOBOLEW, S. L. (Соболев, С. Л.)

[1] О формулах механических кубатур на поверхности сферы, Сиб. матем. журнал **3** (1962), № 5.

[2] Лекции по теории кубатурных формул I, II, Новосиб. ун-т, Новосибирск 1964, 1965.

[3] Введение в теорию кубатурных формул, Наука, Москва 1974.

SOLOTUCHIN, W. G., und G. E. DEJEW (Золотухин, В. Г., и Г. Е. Деев)

[1] Модификация метода Монте-Карло для расчета дифференциального потока нейтронов и γ-квантов, Метод Монте-Карло в проблеме переноса излучений (сб.), Атомиздат, Москва 1967, стр. 53—66.

SOLOTUCHIN, W. G., und S. M. ERMAKOW (Золотухин, В. Г., и С. М. Ермаков)

[1] Применение метода Монте-Карло для расчета защиты от ядерных излучений, Вопросы физики защиты реакторов (сб.), Атомиздат, Москва 1963, стр. 171—182.

SOLOTUCHINA, L. A., K. P. LATYSCHEW und W. N. TSCHUGUJEWA (Золотухина, Л. А., К. П. Латышев и В. Н. Чугуева)

[1] Стохастические модели слоистой среды и вероятностные свойства отраженных волн, распространяющихся в этих средах, Труды МИАН **95** (1968), 42—97.

SPANIER, J., and E. M. GELBARD
[1] Monte Carlo principles and neutron transport problems, Addison-Wesley, Reading (Mass.) 1969.
SRAGOWITSCH, W. G. (Срагович, В. Г.)
[1] Моделирование некоторых классов случайных процессов, Журнал вычислит. матем. и матем. физ. **3** (1963), 586—593.
STAR, O. T., and A. C. C. PAGE
[1] Feasibility financial studies of a port installation, Oper. Res. Quarterly **12** (1961), No 3.
STARTSCHENKO, L. P. (Старченко, Л. П.)
[1] О построении последовательностей совместно нормальных с данной, Изв. АН СССР, сер. матем., **22** (1958), 357—370.
[2] Исправления к работе ,,О построении последовательностей, совместо нормальных с данной'', Изв. АН. СССР, сер. матем., **23** (1959), 635—636.
SZEGÖ, G.
[1] Orthogonal Polynomials, Amer. Math. Soc., Providence (R. I.) 1959.

TAUSSKY, O., and J. TODD
[1] Generation and Testing of pseudo-random numbers, Sympos. on Monte Carlo Methods, ed. H. A. MEYER, J. Wiley & Sons, New York 1956, p. 15—28.
TIPPET, L. H. C.
[1] Random sampling numbers, Tracts for Comp., Cambr., **15** (1927).
TÖRN, A.
[1] Crude Monte Carlo quadrature in infinite variance case and central limit theorem, Nord. Tidskr. Information Behandl. **6** (1966), 339—346.
TSCHENZOW, N. N. (Ченцов, Н. Н.)
[1] О квадратурных формулах для функций бесконечно большого числа переменных, Журнал вычислит. матем. и матем. физ. **1** (1961), 418—424.
[2] Псевдослучайные числа для моделирования марковских цепей, Журнал вычислит. матем. и матем. физ. **7** (1967), 632—643.
TSUDA, T., and T. KIYONO
[1] Application of the Monte Carlo method to systems of nonlinear algebraic equations, Numer. Math. **6** (1964), 59—67.

WAISBORD, E. M., und D. B. JUDIN (Вайсборд, Э. М., и Д. Б. Юдин)
[1] Стохастическая аппроксимация для многоэкстремальных задач в гильбертовом пространстве, ДАН СССР **181** (1968), 1034—1037.
WALLANDER, S. W. (Валландер, С. В.)
[1] Уравнения и постановка задач в аэродинамике разреженных газов, Аэродинамика разреженных газов (сб.), Изд-во ЛГУ **1** (1963), 7—37.
WERSCHIK, A. M., und W. TEMELT (Вершик, А. М., и В. Темельт)
[1] Некоторые вопросы аппрохимации оптимального значения безконечномерных задач линейного программирования, Сиб. матем. журнал **9** (1968), 790—803.
WEYL, H.
[1] Über die Gleichverteilung von Zahlen mod Eins, Math. Ann. **77** (1916), 313—352.
WIENER, N.
[1] Nonlinear problems in random theory, Cambridges (Mass.) Technol. Press/ J. Wiley & Sons, Inc., New York/Chapman & Hall, Ltds., London 1958.
WILDE, D. J.
[1] Optimum seeking methods, Prentice-Hall, Englewood Cliffs (N. J.) 1964.

WILKS, S. S.
[1] Mathematical statistics, J. Wiley & Sons, New York—London 1962.
WLADIMIROW, W. S.
[1] Gleichungen der mathematischen Physik, VEB Deutscher Verlag der Wissenschaften, Berlin 1972 (Übersetzung aus dem Russischen).
WLADIMIROW, W. S., und I. M. SOBOL (Владимиров, В. С., и И. М. Соболь)
[1] Расчет наименьшего характеристического числа уравнения Пайерлса методом Монте-Карло, Вычислит. математика **3** (1958), 130—137.

ZIELIŃSKI, R.
[1] On the Monte-Carlo evaluation of the extremal value of a function, Algoritmy **2** (1965), No 4, 7—13.

Методы Монте-Карло и их применения, Тезисы докладов на 2-м Всесоюзном совещании по методам Монте-Карло 20—25 октября 1969 г., Тбилиси 1969.
Новые идеи в планировании эксперимента (сб.), Наука, Москва 1969.

ERGÄNZUNGSLITERATUR[1]

Monographien zur Monte-Carlo-Methode:

MICHAILOW, G. A. (Михайлов, Г. А.)
[1] Некоторые вопросы теории метода Монте-Карло, Наука, Новосибирск 1974.
POLLJAK, J. G. (Полляк, Ю. Г.)
[1] Вероятностное моделирование на электронных машинах, Сов. радио, Москва 1971.
SOBOL, I. M. (Соболь, И. М.)
[1] Вычислительный метод Монте-Карло, Наука, Москва 1974.
ZIELIŃSKI, R.
[1] Methody Monte Carlo, Warszawa 1970.

Zu Kapitel II:

AHRENS, S. H., and U. DIETER
[1] Computer methods for sampling from the exponential and normal distributions, Comm. Assoc. Comp. Mach. **15** (1972), 873–882.
BELL, I. R.
[1] Algorithm 334: normal random deviates, Comm. Assoc. Comp. Mach. **11** (1968), 498; **12** (1969), 281.
BERESHNOW, W. I., L. K. GORSKI und W. K. KOROLENOK (Бережнов, В. И., Л. К. Горский, и В. К. Короленок)
[1] Алгоритм получения последовательности псевдослучайных чисел с разными законами распределения; Алгоритмы и программы случайного поиска (Сб.), Зинатне, Рига 1969, стр. 335—339.

[1] Die Liste der Ergänzungsliteratur enthält wichtige Arbeiten, die nach 1970 erschienen sind, aber auch einige Arbeiten, die nach Meinung des Autors allgemeine theoretische Situationen gut illustrieren. In fast allen Fällen wurden Titel von Arbeiten angegeben, die nach 1965 erschienen sind. Die Literatur bis 1965 kann man bei Hammersley und Handscomb finden. Besondere Aufmerksamkeit wurde der Beschreibung von Algorithmen gewidmet. Dem Autor ist es klar, daß die angegebene Literatur nicht vollständig ist, und er bittet jene Autoren um Nachsicht, deren Arbeiten nicht aufgenommen werden konnten. Wesentliche Ergänzungen zum vorliegenden Literaturverzeichnis können die von Nance und Claude, aber auch die von Naylor zusammengestellten Literaturverzeichnisse geben. Verweise darauf findet der Leser in dieser Literaturliste.

Burr, W.

[1] A useful approximation of the normal distribution function with application to simulation, Technometrics **9** (1967), 647–651.

Breiman, L.

[1] Probability and stochastic processes, with a view toward applications, Hougton Mifflin Company, Boston (Mass.) 1969.

Dieter, U., and J. Ahrens

[1] A combinatorical method for the Generation of Normally distributed random numbers, Computing **11** (1973), 137–146.

Djadkin, I. G. (Дядькин, И. Г.)

[1] Моделирование случайной энергии гамма кванта, рассеянного в результате Комптон Эффекта, Ж. выч. матем. и матем. физ. **6** (1966), 384—385.

Dobroljubow, L. W., W. A. Likarew und D. I. Popow (Добролюбов, Л. В., В. А. Ликарев и Д. И. Попов)

[1] Датчик нормальных псевдослучайных чисел, основанный на обращении функции распределения Алгоритмы и алгоритмические языки, Вып. 4, ВЦ АН СССР, Москва 1969, стр. 87—94.

Forsythe, G. E.

[1] Von Neumanns Comparison Method for Random sampling from the normal and other distributions, Math. Comp. **26** (1972), 817–826.

Franklin, G. N.

[1] Review of VD Barnett: Random Negative Exponential Deviates, Math. Comp. **20** (1966), 93.

Friedman, I. H.

[1] Random event generation with preferred frequency distributions, J. Comput. Physics **7** (1971), 201–218.

George, R.

[1] Algorithm 200: normal random, Comm Assoc. Comp. Mach. **8** (1965), 559.

German, W. A., und I. M. Sobol (Герман, В. А., и И. М. Соболь)

[1] Об использовании поверхностей постоянной плотности при моделировании многомерных случайных величин, Ж. выч. матем. и матем. физ. **11** (1971), 789—791.

Harding, R. I.

[1] Generation of random digital numbers with specified probability distributions, Radio and electronic engineer **35** (1968), 369–375.

Hurst, R. L., and R. E. Knop

[1] Algorithm 425: generation of random correlated normal variables, Comm. Assoc. Comp. Mach. **15** (1972), 355–357.

Knop, R.

[1] Algorithm 381: random vectors uniform in solid angle, Comm. Assoc. Comp. Mach **13** (1970), 326; **15** (1972), 468.

[2] Algorithm 384: normal random deviates, Comm. Assoc. Comp. Mach. **12** (1969), 281.

Kondjurin, J. N. (Кондюрин, Ю. Н.)

[1] О моделировании случайных величин методом суперпозиций, Ж. выч. матем. и матем. физ. **10** (1970), 262—266.

Král, I.

[1] A note on generation of sequences of pseudorandom numbers with prescribed autocorrelation coefficients, Kybernetika **8** (1972), 485–489,

Ljubimow, Ju. K. (Любимов, Ю. К.)

[1] Получение на ЦВМ дискретных значений стационарного случайного процесса в неравноотстоящих точках, Автомат. и телемех. **26** (1965), 2239.

MARSAGLIA, G.

[1] A general method for producing random variables in a computer, AFIPS Conf. Proc., Vol. 29, Washington 1966, p. 169–173.

MEYER, D. L.

[1] Methods of generating random normal numbers, Educational and psychological measurements **29** (1969), 193–198.

MICHAILOW, G. A. (Михайлов, Г. А.)

[1] Два замечания о моделировании случайних величин, Ж. выч. матем. и матем. физ. **12** (1972), 1353–1357.

NANCE, R. E., and O. J. CLAUDE

[1] Bibliography 29: A Bibliography on random number generation, Comp. Rev. **13** (1972), 495–508.

NEAVE, H. R.

[1] On using the Box-Müller transformation with multiplicative congruential pseudo-random number generators, Appl. Statist. **22** (1973), 92–97.

OLSZEWSKI, T.

[1] Digital computer conversion of pseudorandom numbers with uniform distribution into numbers with a predetermined arbitrary distribution, Prace przemyslowego instytutu telekomunikacji **21** (1971), No 72, 51–55.

OSSOSKOW, G. A. (Ососков, Г. А.)

[1] Генерирование на электронной вычислительной машине случайных чисел с показательным распределением, Ж. выч. матем. и матем. Физ. **7** (1967), 710–713.

PERKO, A.

[1] On the effect of a mixing table on a stationary random sequence, Turun yliopiston julkaisuja Ann. Univ. turkuensis 1971.

PIKE, M. C.

[1] Certification of algorithm 121: Normdeviates, Comm. Assoc. Comp. Mach. **8** (1965), 556.

[2] Algorithm 267: Random normal deviates, Comm. Assoc. Comp. Mach. **8** (1965), 606.

[3] Certification of algorithm 200: Normal random, Comm. Assoc. Comp. Mach. **8** (1965), 556.

PIRANASCHWILI, S. A. (Пиранашвили, З. А.)

[1] Некоторые вопросы статистико-вероятностного моделирования непрерывных случайных процессов, „Вопросы исследования операций" (Сб.), Изд-во Мецниереба, Тбилиси 1966, стр. 53.

POLLJAK, J. G. (Полляк, Ю. Г.)

[1] Об экономических алгоритмах моделирования случайных величин, Ж. выч. матем. и матем. физ. **8** (1968), 710.

[2] Об экономических алгоритмах моделирования случайных величин, Ж. выч. матем. и матем. физ. **6** (1966), 1134.

[3] О моделировании случайных величин, связанных со схемой независимых испытаний, Изв. АН СССР, Тех. Кибернетика № 1 (1969), 210.

[4] Моделирование последовательности неравностоящих по времени выборок из гауссового случайного процесса, Изв. АН СССР, Тех. кибернетика № 1 (1969), 50–56.

RASTRIGIN, L. A. (Растригин, Л. А.)

[1] Дискретное моделирование непрерывных случайных процессов, Докл. у межвузовск. конференции по физ. и матем. моделир. секц. „Матем. вопросы вероятностного и кибернетического моделирования", Москва 1968, стр. 77–82.

Rosin, S. G., K. A. Rosina und L. Ch. Chomski (Розин, С. Г., К. А. Розина и Л. Х. Хомский)

[1] Об одном способе моделирования экспоненциального закона распределения, Ж. выч. матем. и матем. физ. **13** (1973), 503—509.

Rubin, H.

[1] Optimization problems in simulation, Optimizing methods in Statistics, Academic Press, New York—London, 1971, p. 29—32.

Schaffer, H. E.

[1] Algorithm 369: generator of random numbers satisfying the poisson distribution, Comm. Assoc. Comp. Mach. **13** (1970), 49.

Schkurski, B. I. (Шкурский, Б. И.)

[1] Метод математического моделирования двумерных случайных полей, Изв. АН СССР, Тех. кибернетика № 6 (1969), 141.

Schrack, G. F.

[1] Remark on algorithm 381, Comm. Assoc. Comp. Mach. **15** (1972), 468.

Szép, A.

[1] Some remarks on random number transformation, Stud. Sci. Math. Hung. **6** (1971), 393—397.

Zieliński, R.

[1] Generatory liczb losowych, Warszawa 1972.

Zu Kapitel 3:

Alisade, R. I. (Ализаде, Р. И.)

[1] Синтез шарнирно-рычажных механизмов методом статистических испытаний, Машиноведение № 3 (1968), 58—61.

Alsmiller, R. G., T. W. Armstrong and K. C. Chandler

[1] Monte Carlo calculation of high-energy nucleon-meson cascades and comparison with experiment, Nuclear Sci. and Engineering **49** (1972), 82—92.

Barker, J. S. F.

[1] Simulation of genetic systems by automatic digital computers: 3. Selection between alleles at an autosomal locus, Austr. J. Bid. Sci. **11** (1958), 603—612.

[2] Simulation of genetic systems by automatic digital computers: 4. Selection between alleles at a sex-linked locus, Austr. J. Bid. Sci. **11** (1958), 613—625.

Bascharin, G. P., A. D. Charkewitsch und M. A. Schneps (Башарин, Г. П., А. Д. Харкевич и М. А. Шнепс)

[1] Массовое обслуживание в телефонии, Наука, Москва 1968.

Bayes, A. J.

[1] A minimum variance sampling technique for simulation models, J. Assoc. Comp. Mach. **19** (1972), 734—741.

Benson, O.

[1] Simulation of international relations and diplomacy Computer applications in the behavioral, Prentice-Hall, New York 1962.

Burt, J. M. J., and M. B. Garman

[1] Monte Carlo techniques for stochastic pert network analysis INFOR-Canada, J. Operations Res. Inf. Proc. **9** (1971), 248—268.

Burdick, D. S., and T. H. Naylor

[1] Design of computer simulation experiments for industrial systems, Comm. Assoc. Comp. Mach. **9** (1966), 329—339.

Bykow, W. W. (Быков, В. В.)
[1] Цифровое моделирование в статистической радиотехнике, Сов. радио, Москва 1971.

Conway, R. W.
[1] Some tactical problems in digital simulation, Manag. Sci. **10** (1963), 47–61.

Cordaro, M. C., and M. S. Zucker
[1] A method for solving time-dependent electron transport problems, Nuclear Sci. and Engineering **45** (1971), 107–116.

Cushen, W. E.
[1] Operational gaming in industry, Operations Res. **2** (1956), 357–358.

Dal, U., B. Mjurchaug und K. Njugord (Дал, У., Б. Мюрхауг и К. Нюгорд)
[1] СИМУЛА-67 — Универсальный язык программирования, Мир, Москва 1969.

Dean, P., and N. E. Bird
[1] Monte Carlo estimates of critical percolation probabilities, Proc. Camb. Phil. Soc. **63** (1967), 477–479.

Evans, G. W., G. F. Wallace and G. Sutherland
[1] Simulation using digital computers, Prentice-Hall, Englewood Cliffs, New York 1967.

Fetter, R. B., and I. D. Thompson
[1] The simulation of hospital systems, Operations Res. **13** (1965), 689–711.

Fine, G. H., and P. V. McIsaac
[1] Simulation of a time-sharing system, Management Sci. **12** (1966), 180–194.

Fishman, G. S.
[1] Concepts and methods in discrete event digital simulation, John Wiley & Sons, New York 1973.

Fosdick, L. D.
[1] The Monte Carlo method in quantum statistics, SIAM Rev. **10** (1968), 315–328.

Fraser, A. S.
[1] Simulation of genetic systems by automatic digital computers: 1. Introduction Austr. J. Bid. Sci. **10** (1957), 484–491.
[2] Simulation of genetic systems by automatic digital computers: 2. Effects of linkage on rates of advance under selection, Austr. J. Bid. Sci. **10** (1957), 492–499.
[3] Simulation of genetic systems by automatic digital computers: 4. Epistasis, Austr. J. Bid. Sci. **13** (1960), 150–162.
[4] Simulation of genetic systems by automatic digital computers: 7. Effects of reproductive rate and intensity of selection, on genetic structure, Austr. J. Bid. Sci. **13** (1960), 344–350.

Fuksman, A. L., E. D. Charlamowa und N. W. Schmailo (Фуксман, А. Л., Е. Д. Харламова и Н. В. Шмайло)
[1] Ускорение процесса моделирования отказов в больших системах, Автомат. и вычисл. техн. **1** (1969), 63.

Garret, R. E., and A. S. Hall jun.
[1] Optimal synthesis of randomly generated linkages, Trans. ASME, Ser. B, **90** (1968), 475–480.

Gluschkow, W. M., L. A. Kalinitschenko u. a. (Глушков, В. М., Л. А. Калиниченко, Т. П. Марьянович, и др.)
[1] СЛЭНГ — система программирования для моделирования дискретных систем, Изд-во Ин-та Кибернетики АН УССР, Киев 1968.

Gorski, L. K. (Горский, Л. К.)
[1] Статистические алгоритмы исследования надежности, Наука, Москва 1970.
Green, D. H., and M. G. Hartley
[1] The simulation of some simple control policies for a signalized intersection, Operations Res. Quart. **17** (1966), 263–277.
Gussew, W. W. (Гусев, В. В.)
[1] Языки моделирования и некоторые тенденции их развития, Киев 1972.
Harbangh, J. W., and Bonham-Carter
[1] Computer simulation in Geology, John Wiley & Sons, New York 1970.
Hiroshi, Takahashi
[1] Monte Carlo method for geometrical perturbation and its application to the pulsed fast reactor, Nuclear Sci. and Engineering **41** (1970), 259–270.
Huesman, L. R., and R. P. Goldberg
[1] Evaluating computer systems through simulation, Comp. J. **10** (1967), 150–156.
Hutchinson, G. K.
[1] A computation center simulation project, Comm. Assoc. Comp. Mach. **8** (1965), 559–568.
Ionin, G. L., und M. A. Schneps (Ионин, Г. Л., и М. А. Шнепс)
[1] Статистическая эквивалентность оценок по времени и по вызовам в системе массового обслуживания, Латв. матем. ежегодник **2** (1966), 99—102.
Kiviat, P. J., and J. Philip
[1] Development of new digital simulation languages, J. Ind. Eng. **17** (1966), 604–608.
Kleine, H.
[1] A survey of User's views of discrete simulation languages, Simulation **14** (1970), 225–229.
Kofman, A., und R. Krjuon (Кофман, А., и Р. Крюон)
[1] Массовое обслуживание, Мир, Мосвка 1965.
Laski, J. G.
[1] On time structure in (M-C) simulations, Operations Res. Quart. **16** (1965), 329–339.
Levitt, L. B.
[1] The use of self-optimized Exponential Biasing in Obtaining Monte Carlo Estimates of Transmission Probabilities, Nuclear Sci. and Engineering **31** (1968), 500–504.
Linnik, I. J. (Линник, И. Ю.)
[1] Улучшение сходимости метода Монте-Карло в некоторых задачах массового обслуживания, Кибернетика **5** (1973), 129—132.
Maisel, H., and I. Gnugnoli
[1] Simulation of discrete stochastic systems, Sci. Res. Assoc. Inc., Chicago, Palo Alto 1972.
Maislin, I. E. (Майзлин, И. Е.)
[1] Определение временных характеристик системы сетевого планирования со случайными оценками длительностей отдельных работ, Цифровая вычислительная техника и программирование (Сб.), Изд-во Сов. радио, стр. 40.
Margulis C. S., und G. J. Friedman (Маргулис, Х. Ш., и Г. Я. Фридман)
[1] Статистические моделирование некоторого класса сложных систем, Докл. АН СССР **168** (1966), 296—299.
[2] Статистическая модель промышленного предприятия (на примере нефтеперерабатывающего завода), Изв. АН СССР, Техн. кибернетика **3** (1967), 10—16.

MARTIN, F. F.
[1] Computer Modeling and Simulation, John Wiley & Sons, New York 1968.
MCCRACKIN, F. L.
[1] Weighting Methods for Monte Carlo calculation of Polymer configurations, J. Res. Nat. Bur. Stand. B-math. Sci. **76** (1972), No. 3/4, 193–200.
MIHRAM, G. A.
[1] Simulation. Statistical foundations and methodology, Academic Press, New York-London 1972.

NAYLOR, T. H.
[1] Bibliography 19: Simulation and gaming, Comp. Rev. **10** (1969), 61–69.
NAYLOR, T. H., J. L. BALINTEY and D. S. BURDICK
[1] Computer simulation techniques, John Wiley & Sons, New York 1966.
NAYLOR, T. H., D. S. BURDICK and W. E. SASSER
[1] Computer simulation experiments with conomic systems: The problem of experimental design, J. Amer. Statist. Assoc. **62** (1967), 1315–1337.
NAYLOR, T. H., and D. GIANTURCO
[1] Computer simulation in psychiatry, Arch. of General Psychiatry **15** (1966), 293–300.
NAYLOR, T. H., W. H. WALLANCE and W. E. SASSER
[1] A computer simulation model of the textile industry, J. Amer. Statist. Assoc. **62** (1967), 1338–1364.
POLJAK, D. G. (Поляк, Д. Г.)
[1] Оценка точности статистического моделирования систем массового обслуживания, Изв. АН СССР, Техн. кибернетика **1** (1970), 80—88.
[2] Некоторые способы повышения точности статистического моделирования систем массового обслуживания, Изв. АН СССР, Техн. кибернетика **4** (1970), 75—85.
[3] Смешанные оценки для вероятности потери заявки, Проблемы передачи информации **9** (1973), вып. 2, 86—92.
POLJAK, D. G., und E. G. BELOW (Поляк, Д. Г., и Е. Г. Белов)
[1] Сравнение оценок во времени и по вызовам для вероятности потери в многолинейной системе обслуживания с ограниченной очередью, Пробл. передачи информации **8** (1972), вып. 2, 67—82.
POLJAK, D. G., und G. O. ROSENTHAL (Поляк, Д. Г., и Г. О. Розенталь)
[1] О повышении точности статистического моделирования систем массового обслуживания, Автомат. и вычислит. техника **6** (1972), 54—58.
POLLJAK, J. G. (Полляк, Ю. Г.)
[1] Об оптимизации систем по данным статистического моделирования, Вопросы кибернетика и вычислит. матем. (Сб.), Изд-во Фан, Ташкент 1969, стр. 13.
RACHWALSKI, W. M. (Рахвальский, В. М.)
[1] Ускорение процесса моделирования при оценке эффективности и надежности сложных систем методом статистических испытаний, Изв. АН СССР, Тенх. кибернетика **1** (1966), 41.
RAZAFINDRAKOTO, A.
[1] Solutions numériques de l'equation de Poisson á trois variables par la méthode de Monte-Carlo, C. r. Acad. Sci. **A265** (1967), No. 15, 425–428.
SAMALIN, W. M., und G. E. NORMAP (Замалин, В. М., и Г. Э. Нормап)
[1] О методе Монте-Карло в феймановской формулировке квантовой статистики, Ж. выч. матем. и матем. физ. **13** (1973), 408—420.
SCHILDER, M.
[1] A Monte Carlo method for analysing non-linear circuits, Internat. J. Control. **5** (1967), 131–134.

Schneps-Schneppe, M. A. (Шнепс-Шнеппе, М. А.)
[1] Статитисческое моделирование телефонных систем коммутации (Сб.), Наука, Москва 1969, стр. 5—26.
[2] О приложении метода вложенных целей Маркова к моделированию систем массового обслуживания с потерями, Уч. зап. Латв. ун-та им. П. Стучки **47** (1969), 261—274.

Schneps-Schneppe, M. A., und E. A. Inaunieks (Шнепс-Шнеппе, М. А., и Э. А. Инауниекс)
[1] Исследование точности оценок вероятности потерь в однолинейной системе, Латв. матем. ежегодник **6** (1969), 187—206.

Schneps-Schneppe, M. A., und E. I. Schkolny (Шнепс-Шнеппе, М. А., и Е. И. Школьный)
[1] О точности статистического моделирования СМО, описываемых полумарковскими процессами разможения и гибели, Изв. АН СССР, Техн. кибернетика **5** (1973), 60—65.

Taylor, J. G., and J. A. Navarro
[1] Simulation of a court system for the processing criminal cases, Simulation **10** (1968), 241—248.

Thoman, D. R., L. J. Bain and C. E. Antle
[1] Inferences on the Parameters of the Weibull Distribution, Technometrics **11** (1969), 445—460.

Tkatschenko, P. N., L. N. Kuzew, T. A. Mestscherjakow u. a. (Ткаченко, П. Н., Л. Н. Куцев, Т. А. Мещеряков, и др.)
[1] Математические модели боевых действий, Сов. радио, Москва 1969.

Tocher, K. D.
[1] A review of simulation languages, Operations Res. Quart. **16** (1965), 189—217.

Vesely, W. E.
[1] A study of the effects of polarization, electron binding, and Rayleight Scattering in Monte Carlo, gamma-ray transport calculations, Nuclear Sci. and Engineering 1, 2, 3.

Wdowin, A. D., und A. O. Kriwitzki (Вдовин, А. Д., и А. О. Кривицкий)
[1] Исследование методом статистичеких испытаний организации восстановления ненадежных систем, Системы иправления и коммутации, Наука, Москва 1965.
[2] Водноэнергические расчеты методом Монте-Карло, Изд-во Энергия, Москва 1969.

Wood, W. W.
[1] Monte Carlo studies of simple liquid models, In: The Physics of simple Liquids North-Holland Publ. Company, Amsterdam 1968, p. 116—230.

Zieliński, R.
[1] Methody Monte Carlo, Warszawa 1970.

Zu Kapitel 4:

Baldini-Celio, R., B. Ballico-Lay and A. Spano-Mencuccimi
[1] A multistage self-improving Monte-Carlo method, Nucl. Instr. and Methods **72** (1969), 317—320.

Ermakow, S. M., und G. A. Iljuschina (Ермаков, С. М., и Г. А. Илюшина)
[1] О вычислении одного класса интегралов с помощью случайных интерполяционно-квадратурных формул, Методы Монте-Карло в вычислительной математике и математической физике (Сб.), Новосибирск 1974, стр. 67—71.

GRANOWSKI, B. L. (Грановский, Б. Л.)
[1] О дисперсии случайных квадратур гауссовского типа, Исследование операций и статистическое моделирование, I, Изд-во ЛГУ 1972, стр. 53—60.
GRANOWSKI, B. L., und W. G. STOLJAROW (Грановский, Б. Л., и В. Г. Столяров)
[1] К вопросу о допустимости квадратурных формул со случайными узлами, Методы Монте-Карло в вычислительной математике и математической физики (Сб.), Новосибирск 1974, стр. 78—85.
HABER, S.
[1] A combination of Monte-Carlo and classical methods for evaluating multiple integrals, Bull. Amer. Math. Soc. **74** (1968), 683—686.
[2] Stochastic quadrature formulas, Math. Comp. **23** (1969), 751—764.
[3] On a sequence of points of interest for numerical quadrature, J. Res. Nat. Bur. Standards, Ser. B, **70** (1966), 127—136.
HALTON, J. H.
[1] On the relative merits of correlated and importance sampling for Monte Carlo integration, Proc. Cambridge Phil. Soc. **61** (1965) 497—498.
POWELL, M. J. D., and J. SWANN
[1] Weighted uniform sampling a Monte Carlo technique for reducing variance, J. Inst. Math. Appl. **2** (1966), 228—236.
TSUDA, T.
[1] Numerical integration of function of very many variables, Numer. Math. **20** (1973), 377—391.

Zu Kapitel 5:

ANTONOW, G. E., und W. J. KATKOWNIK (Антонов, Г. Е., и В. Я. Катковник)
[1] Метод синтеза одного класса алгоритмов случайного поиска, Автоматика и телемеханика **6** (1971), 154—157.
BHATTACHARYA, P. K.
[1] Estimation of a probability density function and its derivativs, Sankhiya **29** (1967), 373—382.
BOSQ, D.
[1] Complément à deux notes sur l' estimation de la densité et de ses dérivés, C. R. Acad. Sci. Paris **A271** (1970), 45—46.
[2] Estimation non parametrique de la densité et de ses dérivés, C. R. Acad. Sci. Paris **A269** (1969), 1010—1012.
[3] Sur l' estimation d'une densité multivariel par une serie de fonctions orthogonales, C. R. Acad. Sci. Paris **A268** (1969), 555—557.
BRAININA, M. S., W. L. GENEROSOW, W. G. KUSNEZOW und W. A. SOKOWITSCH (Брайнина, М. З., В. Л. Генерозов, В. Г. Кузнецов и В. А. Сокович)
[1] Вычисление производных дозы методом Монте-Карло для оптимизации формы и состава радиационной защиты, Ж. вычислит. матем. и матем. физ. **7** (1967), 953—957.
DMITRIJEW, J. P., und F. P. TARASSENKOW (Дмитриев, Ю. П., и Ф. П. Тарасенко)
[1] Об оценивании функционалов от плотности вероятности и ее производных, Теория вероятностей и ее применения **18** (1973), 662—668.
ERMAKOW, S. M., und W. B. MELAS (Ермаков, С. М., и В. Б. Мелас)
[1] Об одном подходе к планированию эксперимента при нелинейной параметризации, Заводская лаборатория **10** (1973), 1222—1225.

Ermakow, S. M., und E. W. Sedunow (Ермаков, С. М., и Е. В. Седунов)
[1] Рандомизованные планы регрессионных экспериментов с одной свободной точкой, Исследование операций и статистическое моделирование (Сб.), вып. 1, Изд-во ЛГУ 1972, стр. 61—71.

Fedorow, W. A. (Федоров, В. А.)
[1] Теория оптимального эксперимента, Наука, Москва 1971.

Fratila, E.
[1] Monte Carlo method in evaluating a probability density function of random variable, Stud. Univ. Babes-Bolyai Ser. Math.-Mech. **15** (1970), fasc. 2, 27—32.

Katkownik, W. J. (Катковник, В. Я.)
[1] Параметрические операторы усреднения в итерационных алгоритмах решения сложных экстремальных задач, Тезисы докладов VI Всесоюзной конференции по экстремальным задачам, Таллин 1973.

Katkownik, W. J., und O. J. Kultschizki (Катковник, В. Я., и О. Ю. Кульчицкий)
[1] Методы случайного поиска для решения задачи оптимизации управления дискретной системой, Автоматика и вычислит. техника **5** (1972), 49—55.

Kronmal, R., and M. Tarter
[1] The estimation of probability densities and cumulatives by Fourier series method, J. Amer. Stat. Assoc. **63** (1968), 925—952.

Lifšchiz, A. L. (Лившиц, А. Л.)
[1] Использование зависимых испытаний при статистической оптимизации функций, описанных методом Монте-Карло, Вопросы кибернетики и вычислит. матем. (Сб.), Изд-во Фан, Ташкент 1969, стр. 27—30.

Nadaraja, E. A. (Надарая, Э. А.)
[1] Об интегральной среднеквадратической ошибке некоторых непараметрических оценок плотности распределения, VI Международная конференция по теории вероятностей и математической статистике, тезисы докладов, 2, Вильнюс 1973, стр. 109—112.
[2] О сходимости по норме L^2 оценок плотности вероятности, Теория вероятностей и ее применения **18** (1973), 853—856.
[3] О непараметрических оценках плотности вероятности и регрессии, Теория вероятностей и ее применения **10** (1965), 199—203.

Page, E. S.
[1] On Monte Carlo methods in congestion problems I: searching for an optimum in discrete situations, Operation Res. **13** (1965), 291—299.

Parzen, E.
[1] On estimation of a probability density function and mode, Ann. Math. Stat. **33** (1962), 1065—1076.

Poljak, B. T. (Поляк, Б. Т.)
[1] Методы минимизации функций многих переменных (обзор), Экономика и матем. методы **3** (1967), 881.

Renyi, A.
[1] On the theory of random search, Bull. Amer. Math. Soc. **71** (1965), 809—828.

Rosenblatt, M.
[1] Remarks of some nonparametric estimates of a density function, Ann. Math. Stat. **27** (1965), 832—837.

Rubinstein, J. S. (Рубинштейн, Я. С.)
Случайный поиск и процедура Дворецкого, Проблемы случайного поиска I, Зинанте, Рига 1972, стр. 59—64.

SCHWARTZ, S. C.
[1] Estimation of probability density by an orthogonal series, Ann. Math. Stat. **38** (1967), 1261–1265.

SEDUNOW, E. W. (Седунов, Е. В.)
[1] Обобщение задачи Бокса-Дрейпера в планировании регрессионных экспериментов, Заводская лаборатория **3** (1973), 308—314.

SERGIJEWSKI, G. M., und A. P. TER-SAAKJAN (Сергиевский, Г. М., и А. П. Тер-Саакян)
[1] Об использовании факторного эксперимента в задачах многомерной стохастической аппроксимации точки экстремума, Изв. АН СССР, Техн. кибернетика **5** (1970), 134—139.

SOBOL, I. M., und R. B. STATNIKOW (Соболь, И. М., и Р. Б. Статников)
[1] ЛП — поиск и задачи оптимального конструирования, Проблемы случайного поиска I, Зинанте, Рига 1972, стр. 117—136.

STRATONOWITSCH, P. L. (Стратонович, Р. Л.)
[1] Об оптимальных алгоритмах типа стохастической аппроксимации, Изв. АН СССР, Техн. кибернетика **1** (1970), 24—32.

TSCHENZOW, N. N. (Ченцов, Н. Н.)
[1] Оценка неизвестной плотности распределения наблюдений, ДАН СССР **147** (1962), 45—48.
[2] Статистические решающие правила и оптимальные выводы, Наука, Москва 1972.

TSUDA, T., and H. MATSUMOTO
[1] Linear extrapolation of multivariable functions by the Monte Carlo method, Kyoto Univ. Computation Center **1** (1965); J. Assoc. Comput. Mach. **13** (1966), 143–150.

TSUDA, T., and K. ICHIDA
[1] Nonlinear interpolation of multivariable functions by the Monte Carlo method, J. Assoc. Comp. Mach. **17** (1970), 420–425.

WYNN, H. P.
[1] The sequential generation of *D*-optimum experimental designs, Ann. Math. Stat. **41** (1970), 1655–1664.

Zu Kapitel 6:

AMANN, H.
[1] Optimale Anfangsverteilungen bei der Monte-Carlo-Methode mit Informationsspeicherung, Z. Angew. Math. und Mech. **47** (1967), 285–289.

ANOLIK, M. W. (Анолик, М. В.)
[1] Вычисление кратных интегралов, встречающихся в задаче о структуре ударной волны в одноатомном разреженном газе, Методы вычислений (Сб.), **5** Изд-во ЛГУ 1968, стр. 49—60.

ANOLIK, M. W., und R. G. BARANZEW (Анолик, М. В., и Р. Г. Баранцев)
[1] Вторая итерация интегрального кинетического оператора в ударной волне одноатомного газа, Методы вычислений (Сб.), **5** (1968), Изд-во ЛГУ 1968, стр. 43—48.

ANTJUFEJEW, W. S., G. A. MICHAILOW und M. A. NASARALIJEW (Антюфеев, В. С., Г. А. Михайлов и М. А. Назаралиев)
[1] Модификации локальных оценок с учетом осевой симметрии в задачах атмосферной оптики, Вероятностные методы решения задач математической физики (Сб.), Новосибирск 1971, стр. 26—43.

BISCHOFF, F. G., M. L. YEATES and W. E. MOORE

[1] Monte Carlo evaluation of multiple scattering and resolution effects in Double-differential neutron scattering cross-section measurements, Nuclear Sci. and Engineering **48** (1972), 266–280.

BURGART, C. E., and P. N. STEVENS

[1] A general method of importance sampling the angle of scattering in Monte Carlo calculations, Nuclear Sci. and Engineering **42** (1970), 306–323.

CARTER, L. L., E. D. CASHWELL and W. M. TAYLOR

[1] Monte Carlo sampling with continuously varying cross sections along flight paths, Nuclear Sci. and Engineering **48** (1972), 403–411.

CARTER, L. L., and N. J. MCCORMICK

[1] Coupled sampling with the Monte Carlo method in neutron transport calculations, Nuclear Sci. and Engineering **39** (1970), 269–310.

[2] Source convergence in Monte Carlo calculations, Nuclear Sci. and Engineering **36** (1969), 438–441.

CHILTON, A. B.

[1] Linear energy transformation for gammaray Monte Carlo calculations, Nuclear Sci. and Engineering **34** (1968), 328–329.

CHISSAMUTDINOW, A. I. (Хисамутдинов, А. И.)

[1] Оценки с минимальными абсолютными моментами для вычисления методом Монте-Карло суммы ряда Неймана, ДАН СССР **212** (1973), 308—311.

[2] Оценка функционала от решения сопряженного уравнения переноса методом Монте-Карло, Ж. выч. матем. и матем. физ. **8** (1968), 467—471.

[3] ,,Единичный'' класс оценок для вычисления по методу Монте-Карло функционалов от решения интегрального уравнения 2-го рода, Ж. выч. матем. и матем. физ. **10** (1970), 1269—1280.

[4] Оценки ,,единичного'' класса с минимальной дисперсией, Вероятностные методы решения задач математической физики (Сб.), Новосибирск 1971, стр. 184—210.

[5] К уменьшению дисперсии оценки вероятности методом Монте-Карло, Ж. выч. матем. и матем. физ. **10** (1970), 1547—1549.

[6] Выборка по важности в теории переноса излучения, Ж. выч. матем. и матем. физ. **10** (1970), 999—1005.

[7] Об эффективности метода математических ожиданий для задач одного класса, Ж. выч. матем. и матем. физ. **7** (1967), 946—953.

[8] Поведение дисперсии при изменении распределения длины пробега, Ж. выч. матем. и матем. физ. **12** (1972), 257—262.

HANVILAND, J. K.

[1] The solution of two molecular flow problems by the Monte Carlo Method, Methods in Computationel Physics, Vol. 4, Academic Press, New York-London 1965, p. 109–209.

COLEMAN, W. A.

[1] Monte Carlo Calculation of the Effect of Subterrane an Perturbations on Reflected X Rays, Nuclear Sci. and Engineering **46** (1971), 12–21.

COVEYOU, R. R., V. R. CAIN and K. J. YAST

[1] Adjoint and Importance in Monte Carlo Application, Nuclear Sci. and Engineering **27** (1967), 219–234.

DALEZKI, J. L. (Далецкий, Ю. Л.)

[1] Континуальные интегралы, связанные с некоторыми дифференциальными уравнениями и системами, ДАН СССР **137** (1961), 268—271.

DALEZKI, J. L., und L. T. SAPLITNAJA (Далецкий, Ю. Л., и Л. Т. Заплитная)
[1] Интегралы по пространству деревьев, связанные с нелинейными параболическими уравнениями, Укр. матем. ж. **17** (1965), № 5, 110—114.

DENISSIK, S. A., S. N. LEBEDEW und J. G. MALAMA (Денисик, С. А., С. Н. Лебедев и Ю. Г. Малама)
[1] Об одной проверке нелинейной схемы метода Монте-Карло, Ж. выч. матем. и матем. физ. **2** (1971), 783—785.

DJADKIN, I. G., A. T. LISSINENKOW und G. I. PONJATOW (Дядькин, И. Г., А. Т. Лисиненков и Г. И. Понятов)
[1] Об ускорении сходимости метода Монте-Карло при решении задач радиоактивного каротажа, Ж. выч. матем. и матем. физ. **5** (1965), 763—768.

DJADKIN, I. G., und W. N. STARIKOW (Дядькин, И. Г., и В. Н. Стариков)
[1] Об одной возможности экономии машинного времени при решении уравнения Лапласа методом Монте-Карло, Ж. выч. матем. и матем. физ. **5** (1965), 936—938.
[2] Об использовании симметрии и других особенностей нейтронных траекторий для ускорения расчетов методом Монте-Карло, Ж. выч. матем. и матем. физ. **8** (1968), 1001—1012.

ERMAKOW, S. M. (Ермаков, С. М.)
[1] Об аналоге схемы Неймана-Улама в нейлинейном случае, Ж. выч. матем. и матем. физ. **13** (1973), 564—573.
[2] Метод Монте-Карло для итерации нелинейных операторов, ДАН СССР **204** (1972), 271—274.

ERMAKOW, S. M., und W. S. NEFEDOW (Ермаков, С. М., и В. С. Нефедов)
[1] Об оценках суммы ряда Неймана по методу Монте-Карло, ДАН СССР **202** (1972), 27—29.

ERMAKOW, S. M., und P. G. SKWORZOW (Ермаков, С. М., и П. Г. Скворцов)
[1] О достаточных и необходимых условиях несмещенности оценок функционалов, Методы Монте-Карло и их применения (Тезисы докладов третьей Всесоюзной конференции), Новосибирск 1970, 61 стр.

GLADKI, W. S. (Гладкий, В. С.)
[1] Вероятностные вычислительные модели, Наука, Москва 1973.

GLUBIZKI, B. M., und M. W. TANTASCHEW (Глубицкий, Б. М., и М. В. Танташев)
[1] Об органичении дисперсии ,,локальных'' оценок при решении задач переноса излучения методом Монте-Карло, Ж. выч. матем. и матем. физ. **12** (1972), 249—252.

GORELOW, S. L., und M. N. KOGAN (Горелов, С. Л., и М. Н. Коган)
[1] Решение линейных задач динамики разреженного газа методом Монте-Карло, Изв. АН СССР, Мех. жидкости и газа **4** (1969), 136—139.

GRIGORJEW, J. N., M. S. IWANOW und N. I. CHARITONOWA (Григорьев, Ю. Н., М. С. Иванов и Н. И. Харитонова)
[1] Решение задачи о течении Куэтта для кинетического уравнения БГК методом Монте-Карло, Вероятностные методы решения задач математической физики (Сб.), Новосибирск 1971, стр. 69—86.
[2] К вопросу о решении нелинейных кинетических уравнений динамики разреженных газов методом Монте-Карло, Числ. методы механики сплошной среды **2** (1971), № 4, 101—107.

HARTMANN, W.
[1] Über eine Erweiterung der Monte-Carlo-Methode zur Lösung linearer Randwertprobleme zweiter und dritter Art, Z. Angew. Math. und Mech. **52** (1972), 471—477.

HOSHINO, S., and K. ICHIDA
[1] Solution by a modified random walk, Numer. Math. **18** (1971), 61—72.

JASINSKI, F. N. (Ясинский, Ф. Н.)

[1] Применение метода Монте-Карло к нелинейным аэродинамическим проблемам, Изв. АН СССР, Сер. мех. жидк. и газа **6** (1968), 69—73.

JELEPOW, B. S., und G. A. MICHAILOW (Елепов, Б. С., и Г. А. Михайлов)

[1] Алгоритм ,,блуждания по сферам" для уравнения $\Delta u - cu = -g$, ДАН СССР **212** (1973), 15—18.

JELEPOW, B. S., G. A. MICHAILOW und G. R. STEFANJUK (Елепов, Б. С., Г. А. Михайлов и Г. Р. Стефанюк)

[1] Алгоритмы метода Монте-Карло для решения задачи Дирихле в сложных областях с одновременной оценкой градиента потенциала, Тр III Казахстан. межвуз. научн. конф. по матем. и мех. 1967, Алма-Ата 1970, стр. 151—153.

KARCHER, R. H., R. C. ERDMANN and O. C. BALDONADO

[1] The application of the Track-Lenght Distribution Biasing in Monte Carlo Deep-Penetration Calculations, Nuclear Sci. and Engineering **31** (1968), 492—499.

KARGIN, B. A., und G. A. MICHAILOW (Каргин, Б. А., и Г. А. Михайлов)

[1] Исследование эффективности использования асимтотических решений в расчетах по методу Монте-Карло, Ж. выч. матем. и матем. физ. **12** (1972), 150—158.

KOGAN, M. N. (Коган, М. Н.)

[1] Динамики разреженных газов, Наука, Москва 1967.

KONDJURIN, J. N. (Кондюрин, Ю. Н.)

[1] Некоторые способы повышения точности метода Монте-Карло с применением к расчету вероятностных характеристик нелинейных систем автоматического регулирования, Ж. выч. матем. и матем. физ. **8** (1968), 1167—1173.

[2] Некоторые алгоритмы метода Монте-Карло для расчета функционалов, определяемых траекториями нелинейной системы стохастических дифференциальных уравнений, Ж. выч. матем. и матем. физ. **10** (1970), 1375 до 1384.

KSCHWENDT, H.

[1] Convergence limits in the Monte Carlo theory of integral Equations, Numer. Math. **11** (1968), 307—314.

LANORE, J. M.

[1] Weighting and Biasing of a Monte Carlo Calculation for very deep Penetration of Radiation, Nuclear Sci. and Engineering **45** (1971), 66—77.

LEVITT, L. V., and J. SPANCER

[1] A new Non Multigroup Adjoint Monte Carlo Technique, Nuclear Sci. and Engineering **37** (1969), 278—287.

MACMILLAN, D. B.

[1] Optimization of Importance-Sampling parameters in Monte Carlo, Nucl. Sci. and Engineering **48** (1972), 219.

MENDELSON, M. R.

[1] Monte Carlo Criticality Calculations for Thermal Reactors, Nuclear Sci. and Engineering **32** (1968), 319—331.

MICHAILOW, G. A. (Михайлов, Г. А.)

[1] Некоторые замечания об исполнении сопряженных решений для улучшения алгоритмов метода Монте-Карло, Исследование операций и статистическое моделирование **1** (Сб.), Изд-во ЛГУ 1972, стр. 112—118.

[2] Новый алгоритм метода Монте-Карло для оценки максимального собственного значения интегрального оператора, ДАН СССР **191** (1970), 993—996.

[3] Метод моделирования длины свободного пробега частицы, Атомная энергия **28** (1970), № 2, 175.

[4] О сочетании метода конечных сумм с методом Монте-Карло для решения интегральных уравнений 2-го рода, Матем. заметки **9** (1971), 425—434.

[5] Расчеты критических систем методом Монте-Карло, Ж. выч. матем. и матем. физ. **6** (1966), 71—80.

[6] О расчетах возмущений ядерных реакторов методом Монте-Карло, Ж. выч. матем. и матем. физ. **6** (1966), 380—384.

[7] Некоторые замечания об использовании сопряженных решений для улучшения алгоритмов метода Монте-Карло, Исследование операций и статистическое моделирование **1** (Сб.), Изд-во ЛГУ, стр. 112—119.

[8] Вычисление методом Монте-Карло производных функционалов от решения уравнения переноса по параметрам систем, Ж. выч. матем. и матем. физ. **7** (1967), 915—919.

[9] Использование приближенных решений сопряженной задачи для улучшения алгоритмов метода Монте-Карло, Ж. выч. матем. и матем. физ. **9** (1969), 1145—1152.

[10] Оценка некоторых нелинейных функционалов и приближенный расчет групповых констант теории переноса методом Монте-Карло, Ж. выч. матем. и матем. физ. **8** (1968), 590—599.

MICHAILOW, G. A., und M. A. NASARALIJEW (Михайлов, Г. А., и М. А. Назаралиев)

[1] Оптимизация оценки функциональных зависимостей методом Монте-Карло, Ж. выч. матем. и матем. физ. **10** (1970), 734—740.

MISCHUSTIN, B. A. (Мишустин, Б. А.)

[1] О решении задачи Дирихле для уравнения Лапласа методом статистических испытаний, Ж. выч. матем. и матем. физ. **7** (1967), 1179—1187.

NEKRUTKIN, W. W. (Некруткин, В. В.)

[1] Вычисление интегралов по постранству деревьев методом Монте-Карло, в: Методы Монте-Карло в вычислительной математике и математической физике, Новосибирск 1974, стр. 94—102.

NISSEN-MEYER, S.

[1] Some theorems in the problem of reversibility of mathematical models, Rept. Appl. Math. Univ. Bergen **34** (1972), 25.

NORDSIECK, A., und B. L. HICKS

[1] Monte Carlo evaluation of the Boltzmann collision integral, Rarefied Gas Dynamics, Vol. 1, ed. by C. L. BRUNDIN, Academic Press, New York-London **1967**, 695—710.

PERLMUTTER, (Перлмуттер, 000)

[1] Решение задач о течении Куэтта и о теплопередаче между параллельными пластинами в разреженном газе методом Монте-Карло, Выч. методы в динамике разреженных газов (Сб.), Мир, Москва 1969, стр. 116—139.

STEINBERG, H. A., and M. H. KALOS

[1] Bounded Estimators for Flux at a Point in Monte Carlo, Nuclear Sci. and Engineering **44** (1971), 406—412.

STARGUT, J., und J. ROZEWICZ

[1] Zastosowanie metody Monte Carlo do tzezeciego zagadnienia brzegowego ustatonego przewodzenia ciepla, Zesz. nauk Politechn śląskiej **168** (1966), 3—30.

WESSELOWSKAJA, A. S. (Веселовская, А. З.)

[1] О решении первой краевой задачи для волнового уравнения методом Монте-Карло, в: Методы Монте-Карло в вычислительной математике и математической физике, Новосибирск 1974, стр. 128—135.

Wlassow, W. I. (Власов, В. И.)
[1] Улучшение метода статистических испытаний (Монте-Карло) для расчета течений разреженных газов, ДАН СССР, **167** (1966), 1016—1018.
[2] Расчет методом Монте-Карло потока тепла между параллельными пластинами в разреженном газе, Ученые записки ЦАГИ **1** (1970), № 4, 46—51.
Wlassow, W. I., und J. I. Chlopkow (Власов, В. И., и Ю. И. Хлопков)
[1] Вариант метода Монте-Карло для решения линейных задач динамики разреженного газа, Ж. выч. матем. и матем. физ. **13** (1973), 1075—1080

Zu Kapitel 7:

Ahrens, J., U. Dieter and A. Grube
[1] Pseudo-random numbers: A new proposal for the choise of multiplicators, Computing **6** (1970), 121—138.
Antipow, M. W. (Антипов, М. В.)
[1] Оптимальный производящий множитель мултипликативного датчика, Вероятностные методы решения задач матем. физ. (Сб.), Новосибирск 1971, стр. 7—25.
Beyer, W. A., R. B. Roof and D. Williamson
[1] The lattice structure of multiplicative congruental pseudo-random vectors, Math. Comp. **25** (1971), No. 114, 345—363.
Butler, E. L.
[1] Algorithm 370: general random number generator, Comm. Assoc. Comp. Mach. **13** (1970), 49—52.
Cenacci, G., and A. de Matteis
[1] Quasi-random sequences by Power Residues, Numer. Math. **20** (1972), 54—63.
Coveyou, R. R.
[1] Serial correlation in the generation of pseudo-random numbers, J. Assoc. Comp. Mach. **7** (1960), 72—74.
Dieter, U.
[1] Statistical interdependence of pseudo-random numbers generated by the linear congruential method, Proc. Symp. on Applic. of Number Theory to numerical Analysis (Montreal 1971), Academic Press, New York 1972, p. 287—317.
[2] Pseudo-random numbers: the exact distribution of pairs, Math. Comput. **25** (1971), 855—883.
Dieter, U., and J. Ahrens
[1] An exact determination of serial correlations of pseudo-random numbers, Numer. Math. **19** (1971), 101—123.
Djadkin, I. G., N. A. Riswanowa und N. A. Pogorelowa (Дядькин, И. Г., Н. А. Ризванова и Н. А. Погорелова)
[1] Об алгоритмах случайных чисел для машины ,,Раздан-2'', Ученые записки Башкирского университета, сер. матем. **20** (1965), 132.
Ermakow, S. M. (Ермаков, С. М.)
[1] Замечание о псевдослучайных последовательностях, Ж. выч. матем. и матем. физ. **12** (1972), 1077—1082.
Fedortschenko, W. E. (Федорченко, В. Е.)
[1] Моделирование равномерных псевдослучайных чисел на машине ,,Сетунь'', Сер. Матем. обслуживание машины ,,Сетунь'', вып. 15, Издво МГУ 1966.
Felder, H.
[1] The gpss/360 random number generator second conference on application of simulation, Inst. Electrical and Electronics Engineers (1968), 21—28.

GROSENBAUGH, L. R.
[1] More on fortran random generators (short communication), Comm. Assoc. Comp. Mach. **12** (1969), 639.
GUSTAVSON, F. G., and W. LINIGER
[1] A fast random number generator with good statistical properties, Computing **6** (1970), 221–226.
HANSSON, L.
[1] Remark on algorithm 266 [G5]. Pseudo-random numbers, Comm. Assoc. Comp. Mach. **9** (1966), 687.
HEMMERLE, W. I.
[1] Generating pseudo-random numbers on a two's complement machine such as the IBM 360, Comm. Assoc. Comp. Mach. **12** (1969), 382–383.
HSU, L. C.
[1] A general approximation method of evaluating multiple integrals, Tohoku Math. J. **9** (1957), 45–55.
HUFNAGEL, R. E., and E. L. KERR
[1] A simple algorithm for fast real-time generation of pseudo-random poisson integers with rapidly varying means, Proc. IEEE **57** (1969), 2088.
HWANG, E. K., and S. LIN
[1] On generating a random sequences, J. Appl. Probab. **8** (1971), 366–373.
KONJUCHOWSKI, W. W. (Конюховский, В. В.)
[1] Моделирование равномерного распределения, Рабоа ВЦ МГУ (Сб.), вып. 18 (1972), стр. 138—154.
[2] Арифметическое моделирование процесса Пуассона, Рабоа ВЦ МГУ (Сб.), вып. **18** (1972), 125—137.
KUTSCHERENKO, W. E. (Кучеренко, В. Е.)
[1] Датчики псевдослучайных чисел для ЭВМ ,,Сетунь'', Труды ВНИКА нефт. и. газ, вып. **2**, Автоматизация технологических процессов, (Сб.) Изд-во Недра, 1968, 186–189.
LEWIS, P. A. W., A. S. GOODMAN and J. M. MILLER
[1] A pseudo-random number generator for the system 360, IBM system J. **8** (1969), 136–146.
NANCE, R. E., and O. J. CLAUDE
[1] Bibliography 29. A Bibliography on Random number generation, Comp. Rev. **13** (1972), 495–508.
NIDERREITER, H.
[1] On the distribution of pseudo-random numbers generated by the Linear Congruentiel Method, Math. Comp. **26** (1972), 793–795.
NOWAKOWSKI, A.
[1] Generator of pseudo-random numbers for the odra 1204 computer, Prace przemyslowego instytute telekomunikacji **21** (1971), No. 72, 57–58.
PAYNE, W. H.
[1] Fortran Tausworthe pseudorandom number generator (short communication), Comm. Assoc. Comp. Mach. **13** (1970), 57.
PAYNE, W. H. I. R. RABUNG and T. P. BOGYO
[1] Coding the Lehmer Pseudorandom Number Generator, Comm. Assoc. Comp. Mach. **12** (1969), 85.
PIKE, T. C., and I. D. HILL
[1] Algorithm 266: Pseudo-random numbers, Comm. Assoc. Comp. Mach. **8** (1965), 605–606.
[2] Algorithm 266: Pseudo-random numbers, Comm. Assoc. Comp. Mach. **9** (1966), 687.

POLLJAK, J. G. (Полляк, Ю. Г.)
[1] Об анализе псевдослучайных чисел, Автомат. и выч. техника **5** (1968), 31—35.

PROLL, L. G.
[1] Remark on algorithm 370, Comm. Assoc. Comp. Mach. **15** (1972), 467—468.

RADER, C. M.
[1] A fast method of generating digital random numbers, Bell systems technical J. **49** (1970), 2303—2310.

SERAPHIN, D. S.
[1] A fast random number generator for IBM 360, Comm. Assoc. Comp. Mach. **12** (1969), 695.

SOBOL, I. M. (Соболь, И. М.)
[1] Три замечания о проверке псевдослучайных чисел, в: Методы Монте-Карло и их применения (Тезисы докладов на 2-ом Всесоюзном совещании по методам Монте-Карло), Изд-во ИК АН ГССР, Тбилиси 1969, стр. 183.

STONEHAM, R. G.
[1] On a new class of multiplicative pseudo-random number generators, Nord. tidskr. for informationsbehandling **10** (1970), 481—500.

STROME, W. M.
[1] Algorithm 294: uniform random numbers, Comm. Assoc. Comp. Mach. **10** (1967), 40.

SWICH, C. S.
[1] Multiplicative pseudo-random number generators with prime modulus, J. Assoc. Comp. Mach. **18** (1971), 586—593.

TOOTILL, I. P. R., W. D. ROBINSON and A. G. ADAMS
[1] The runs up-and-down performance of Tausworthe pseudo-random number generators, J. Assoc. Comp. Mach. **18** (1971), 381—399.

VINCENT, C. H.
[1] Precautions for accuracy in the generation of truly random binary numbers, J. Physics **4** (1971), 825—828.

WORONZOW, J. W., und J. G. POLLJAK (Воронцов, Ю. В., и Ю. Г. Полляк)
[1] Об использовании квазислучайных последовательностей при прямом вероятностном моделировании систем, Автомат. и выч. техника **6** (1971), 23—27.

SACHVERZEICHNIS





